Author:
Walker, Paul

Title:
Mechanisms for the Generation of Structural Diversity in Polyketide Biosynthesis

Mechanisms for the Generation of Structural Diversity in Polyketide Biosynthesis

Paul David Walker

A thesis submitted to the University of Bristol as part of the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

University of Bristol

School of Chemistry

Cantock's Close

Bristol, BS8 1TS

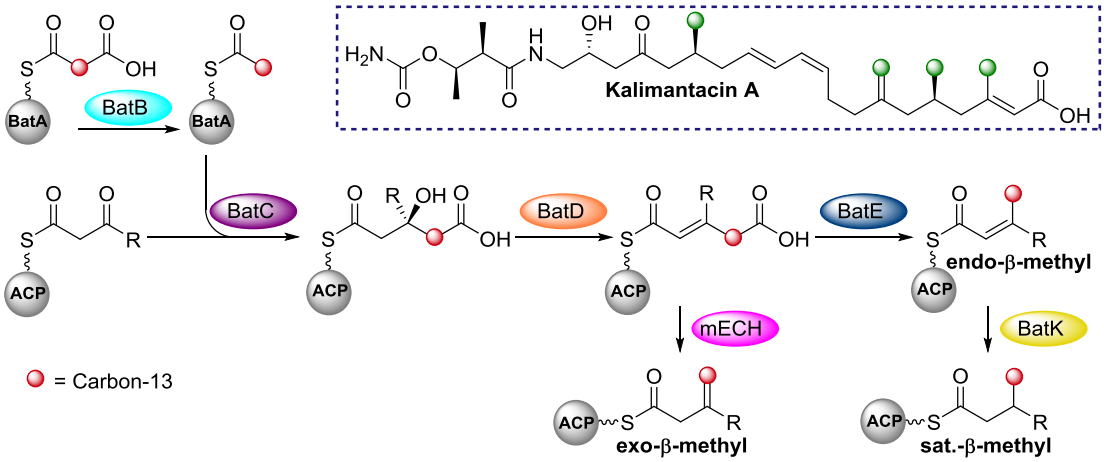
January 2019

Abstract

Kalimantacin β-Branching Pathway

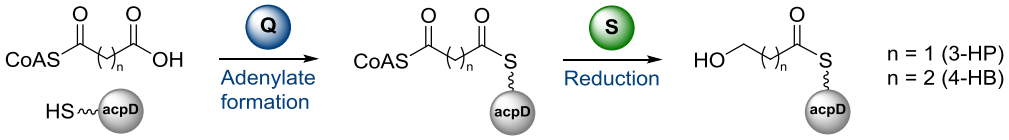
Polyketide synthases encode a remarkable number of enzymes to catalyse chain extension, reductive processing and tailoring reactions leading to structural diversity in the polyketide products. Interaction of a modular ACP-bound substrate with a *trans*-acting HCS cassette leads to β-branching, as exemplified by kalimantacin A which contains four β-branches with an endo or exo double bond or a saturated methyl group (green dots). This diversity of β-branches, incorporated by a single HCS cassette, provides a fascinating case study to probe the mechanism and selectivity for consecutive β-branch incorporations.

The HCS cassette, modular ACPs and tailoring enzymes from the kalimantacin gene cluster were cloned, expressed and purified. *In vitro* reconstitution of the β-branching pathway demonstrated the formation of an *endo*-β-methyl branch by the HCS cassette. The α,β-unsaturation was rigorously assigned by NMR experiments using pantetheine- and ACP-bound intermediates. Subsequent reduction (BatK) of this substrate resulted in the formation of a saturated-β-methyl branch. A previously unassigned modular ECH domain was identified, cloned and expressed as a single domain (mECH) or ACP4-mECH di-domain (4M). An NMR assay utilising a single carbon-13 label incorporated into key biosynthetic mimics was developed and definitively showed the formation of an *exo*-β-methyl branch *via* mECH-catalysed decarboxylation.



Mupirocin/Thiomarinol Starter Unit Generation

Previous stable isotope labelling studies carried out on the mupirocin (*P. fluorescens*) and thiomarinol (*Pseudoalteromonas* sp.) producing strains was the basis for a hypothesis for the formation of a 3- or 4-carbon starter unit in the biosynthesis of the fatty acid moiety of both compounds. The cassette of enzymes hypothesised to be responsible for the generation of a 3-hydroxypropionate (3-HP) and 4-hydroxybutyrate (4-HB) starter unit consists of an ACP (M/TacpD), an adenylation domain (MupQ/TmlQ) and a reductase (MupS/TmlS). Each of these proteins were expressed and purified, and the proposed pathway was then reconstituted *in vitro* which conclusively demonstrated 3-HP and 4-HB formation.



Acknowledgements

I feel very fortunate to have spent four years in Bristol, working on a project I loved and surrounded by talented people. I would like to thank Matt and Chris for their guidance and support during this project. You gave me freedom to think and develop my own ideas, but you were always there to help during the challenging times. Many an hour was spent pondering the different possibilities for β-branching either in your offices or in the NMR room. You allowed me to travel to conferences (Belgium, Jamaica) and for hockey (too many to mention), and I am grateful for those opportunities and your support. Thank you also to Paul Race and Tom Simpson for your input and guidance.

I was very lucky to be able to work between the Crump and Willis research groups and with so many different people. I spent an incredible time with undergraduates, PhD students and post-docs from all over the world with a huge range of talents and experience. It is the people you spend time with in the lab that shape your experience and I am so pleased I had the opportunity to work in such a diverse and rewarding place. In particular I would like to single a few people out. David and I spent the entirety of our PhD working together, from PACT, through my early synthetic efforts, protein production next door and then returning to ruin your peace in the final few months when you thought you had two fumehoods. You were a constant during my PhD, a great friend and support in the lab, and a keen volunteer for running on the Downs. Jonny also made the cut for Monday night running and I'm grateful for the many messages of support whilst sat on the windowsill, surrounded by tomatoes, in those last few months. Angus, without your help at the end there is no way the project would have ended successfully, so thank you! You were a delight to supervise (sometimes), a joy to work with (maybe) and overall a great character to have in the lab (apparently). To my desk neighbour, Luoyi, I apologise for my prying eyes, endless questions and teaching you lesser-known English phrases. It was a pleasure to work and share ideas with you and I sincerely wish you a very successful future. Edith, you seemed to occupy all the desks close to mine at some point. You are a kind and caring person who made me smile and laugh through the good times and the bad. To Nick, thank you for the example you set in the lab, the dedication

to your work and a great set of office games. Chickenlegs (Ash), thank you for the time we spent together. From the bluntest of introductions on my first day in the biosuite you became a great friend! I miss those long days of working on the FPLCs and moaning about…everything! Dr Williams, thank you for teaching me all I know about protein and listening to my endless ramblings trying to work out what to do next. Mr Rowe (snowflake), another undergrad I couldn't get rid of! Thank you for your contribution to the work in this thesis, in particular your knowledge of upfield shifted methyls. To Kun, Catherine, Abi, Joe, Jawaher, Sbu, Beth, Hannah, Song, Kate, Dao, Katherine, Alex, Dan, Claudio, Emma, Hanim, Lisa, Freya, Hornsby, Tom A, Burnham, Alice, Erik, Nahida, Goodwin, Lina and Dr Knowles…thank you!

I would like to thank the Bristol Chemical Synthesis CDT and EPSRC for funding my research. Thank you to Kevin Booker-Milburn and the management team for the work they do to run the programme. Also, to the 2014 CDT cohort, who I worked so closely with during PACT before we started our research projects. I fondly remember the time we spent together in North Wales and walking up Snowdon before project allocation.

Thank you to those non-university friends we made in Bristol for making our time here so special. Rachel, thank you for letting me stay with you when I came back whilst writing.

Finally, I would like to thank my amazing wife, Alison. 2014 was a whirlwind year when we got married, honeymooned in New Zealand and bought our first house. We ended it by moving to Bristol to start my PhD and your theological training. It was the start of the most incredible four years doing the things we love, and I know we both treasure that time. Thank you for your patience, support and the many Bravas nights. I love you!

Author's Declaration

The work described in this thesis was carried out in the School of Chemistry, University of Bristol under the supervision of Professor M.P. Crump and Professor C. L. Willis between September 2014 and September 2018. The work is original, except where indicated by reference in the text, and has not been submitted for any other degrees. The views expressed in the thesis are those of the author and in no way represent those of the University of Bristol.

Paul David Walker

March 2019

Abbreviations

$[\alpha]_D$	Specific Optical Rotation
3-HP	3-Hydroxypropionate
4-HB	4-Hydroxybutyrate
4M	ACP4-mECH
A	Adenylation domain
Ac	Acetyl
Acac	Acetoacetyl
ACP	Acyl Carrier Protein
ACP_A	Acceptor ACP
ACP_D	Donor ACP
AcpS	Acyl Carrier Protein Synthase
aq.	Aqueous
Ar	Aryl
AT	Acyl Transferase
B	Branching domain
BLAST	Basic Local Alignment Search Tool
br	Broad
c	Concentration
C	Condensation domain
CD	Circular Dichroism
CDI	Carbonyl Diimidazole
CID	Collision Induced Dissociation
CoA	Coenzyme A
CS	Crotonase Superfamily
d	Doublet
Da	Dalton
DEBS	6-Deoxyerythronolide B Synthase
DH	Dehydratase
DIBAL-H	Diisobutylaluminum Hydride
DIPEA	Diisopropylethylamine
DKP	Diketopiperazine
DMAP	4-Dimethylaminopyridine

DMAPP	Dimethylallyl Diphosphate
ECH	Enoyl-CoA Hydratase
EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimde
ER	Enoyl Reductase
ESI-MS	Electrospray Ionisation-Mass Spectrometry
FPLC	Fast Protein Liquid Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
HCS cassette	3-Hydroxy-3-methylglutaryl CoA-synthase Cassette
HMG	3-Hydroxy-3-methylglutaryl
HMGCS	3-Hydroxy-3-methylglutaryl CoA-synthase
HMGR	3-Hydroxy-3-methylglutaryl CoA-reductase
HMGS	3-Hydroxy-3-methylglutaryl synthase
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Correlation
Hz	Hertz
Imid.	Imidazole
IPI	Isopentyl Pyrophosphate Isomerase
IPP	Isopentyl Diphosphate
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
IR	Infrared
J	Coupling Constant
KR	Ketoreductase
KS	Ketosynthase
KS^0	Ketosynthase (lacking cysteine for condensation)
LC/MS	Liquid Chromatography/Mass Spectrometry
LDA	Lithium Diisopropylamide
Lit.	Literature
m	Multiplet
M	Molar
m.p.	Melting Point
m/z	Mass-to-charge Ratio
MDD	Mevalonate Diphosphate Decarboxylase

mECH	Modular Enoyl-CoA Hydratase
MG	3-Methylglutaconyl
MHz	Megahertz
MK	Mevalonate Kinase
MS	Mass Spectrometry
MT	Methyl Transferase
NaHMDS	Sodium Hexamethyldisilazide
NCE	New Chemical Entity
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NRPS	Non-ribosomal Peptide Synthase
OAH	Oxyanion Hole
OD	Optical Density
ORF	Open Reading Frame
Pant	Pantetheine
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
petrol	Petroleum Ether 40-60 °C
PK	Polyketide
PKS	Polyketide Synthase
PMK	Phosphomevalonate Kinase
Ppant	Phosphopantetheine
ppm	Parts per Million
PPTase	Phosphopantetheinyl Transferase
PropAc	Propionylacetate
q	Quartet
R_f	Retention Factor
s	Singlet
SAM	*S*-adenosylmethionine
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis
Sfp	Surfactin Phosphopantetheinyl Transferase

SNAC	*N*-acetylcysteamine
t	Triplet
TBS	*tert*-butyldimethylsilyl
TECP	Tris(2-carboxyethyl)phosphine
TE	Thioesterase
tert	Tertiary
TEV	Tobacco Etch Virus
TLC	Thin Layer Chromatography
TMS	Trimethylsilyl
TOCSY	Total Correlation Spectroscopy
δ	Chemical Shift

Table of Contents

1 Introduction 1

1.1 Antimicrobial Resistance 1

1.2 Polyketides 2

1.3 Polyketide Biosynthesis 3

1.4 β-Branching pathway 7

1.5 Kalimantacin 20

1.6 Project Aims 26

2 Kalimantacin β-Branching Pathway 27

2.1 In Vitro Reconstitution of the HCS Cassette 27

2.1.1 Identification of a missing mECH domain 27

2.1.2 Purification of HCS cassette and ACPs 32

2.1.3 Purification of tailoring domains 39

2.1.4 Synthesis of pantetheines 41

2.1.5 Mass spectrometry assay for ACP4-6 45

2.1.6 Unsaturated β-methyl branches 48

2.1.7 Saturated-β-methyl branches 53

2.2 NMR Characterisation of β-Branching 56

2.2.1 Pantetheine NMR experiments 56

2.2.2 4M expression and structural analysis 65

2.2.3 Synthesis of [^{13}C]-labelled materials 68

2.2.4 ^{13}C NMR assay for ACP4 73

2.2.5 ^{13}C NMR assay for 4M 83

2.2.6 Bioinformatic analysis of ECH_2 domains 88

2.2.7 Mechanisms for β-branch selectivity 91

2.3 Kalimantacin Isolation, Analysis and Synthetic Studies 93

2.3.1 Natural product isolation and stability 93

2.3.2 Synthetic studies 97

2.4 Conclusions and Future Work 102

2.4.1 Conclusions 102

2.4.2 Future Work 104

3 Mupirocin/Thiomarinol Starter Unit Generation 105

3.1 Introduction to mupirocin/thiomarinol 105

3.2 Results and Discussion 115

3.2.1 Protein expression and purification 115

3.2.2 In vitro starter unit generation 117

3.3 Conclusions and Future Work 122

3.3.1 Conclusions 122

3.3.2 Future Work 122

4 References 124

5 Experimental 138

5.1 Materials and Methods 138

5.2 Compound Characterisation 152

6 NMR Spectra 175

7 Appendices 209

7.1 Sequences 209

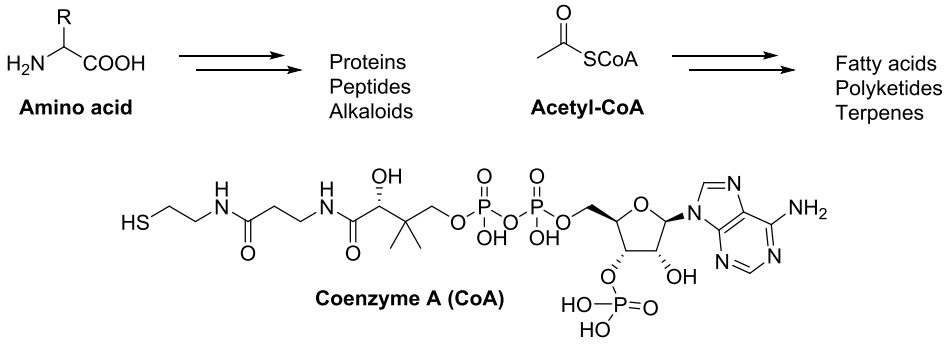
7.2 Sequence Alignments 214

1 Introduction

1.1 Antimicrobial Resistance

Antimicrobial resistance (AMR) is a global threat to healthcare, economics and security according to the 2016 Review on Antimicrobial Resistance.[1] The authors predicted the mortality rate from AMR could reach 10 million people per year by 2050, an increase from the current rate of 700,000. Several key recommendations were made including increasing public awareness, reducing unnecessary use of antimicrobials, improving diagnostics and increasing the number of effective antimicrobials available.

Following the discovery of penicillin in 1928, a golden age of antibiotic discovery followed from 1940 to 1960. A medicinal chemistry driven optimisation of existing chemical scaffolds has ensued since; however, no new classes of antibiotics were introduced until 2000.[2,3] Between 1981 and 2010, 69% of all antimicrobial new chemical entities (NCE) were naturally derived or inspired. At a time when resistance is on the rise, antibiotic discovery and development is in decline.[4,5] Utilising modern synthetic chemistry is one way to develop more effective antibiotics as demonstrated by Boger *et al.* who have reported potent derivatives of the clinically relevant antibiotic vancomycin.[6] Understanding and rational engineering of biosynthetic pathways is another important approach to developing more effective antimicrobials.[7,8] In addition to new and more potent compounds, Peschel *et al.* presented the importance of selective "decolonisation agents" in the control of antibiotic-resistant populations in the human microbiome.[9]



Scheme 1 Biological macromolecules are derived from starter units including amino acids and coenzyme A thioesters.

A range of simple building blocks are constructed into the biomolecules that sustain life in all organisms. This pool of simple, organic molecules is shared between primary and secondary metabolism. Primary metabolites are those that are essential to growth and

reproduction of the organism, whereas secondary metabolites are non-essential but produced for protection or to aid survival. Amino acids are the building blocks utilised in the production of proteins, peptides and alkaloids, whilst acetyl-CoA is used in the biosynthesis of fatty acids, terpenes and polyketides (Scheme 1).

1.2 Polyketides

Polyketides are a diverse range of secondary metabolites produced by bacteria, fungi and marine organisms. Many polyketides have found use in a clinical setting, displaying a wide array of biological activity, including antibiotic, anti-cancer, immunosuppressants and anti-parasitic (Figure 1).[10,11]

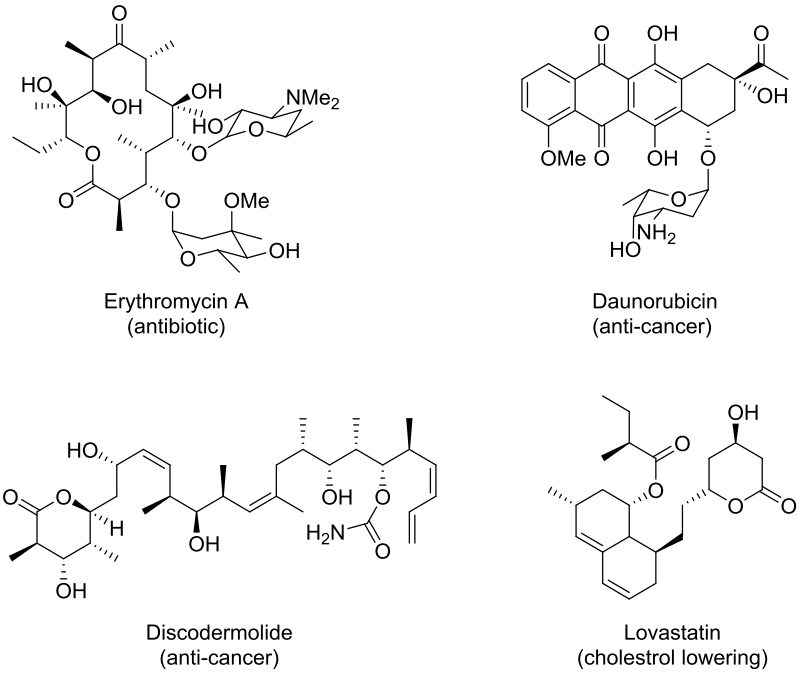
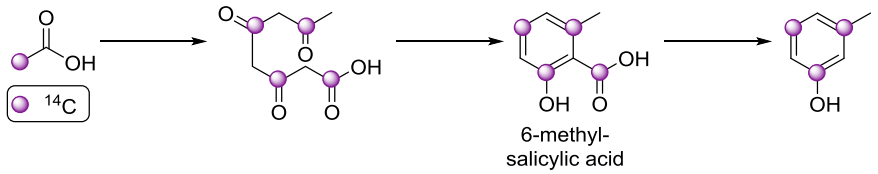


Figure 1 Examples of polyketides and their biological activity.

This class of natural product was first reported by Collie in 1893,[12] and the same author proposed in 1907 the origin of polyphenols to be a repeating -CH_2CO- "ketide" building block.[13] It wasn't until the 1950s, however, that Birch speculated that 6-methyl-salicylic acid was produced from a polyketone that underwent condensation to form the aromatic compound (Scheme 2).[14] This was elegantly demonstrated by the incorporation of ^{14}C by the producing organism and subsequent analysis by chemical degradation. Since these early experiments, isotopic labelling has been a critical tool in elucidating the structures and understanding the biosynthetic pathways of many polyketides.[15] This includes the feeding of singly or doubly labelled acetate, or other biosynthetic precursors, and analysing the NMR spectra of the products for signal enhancement.[16]



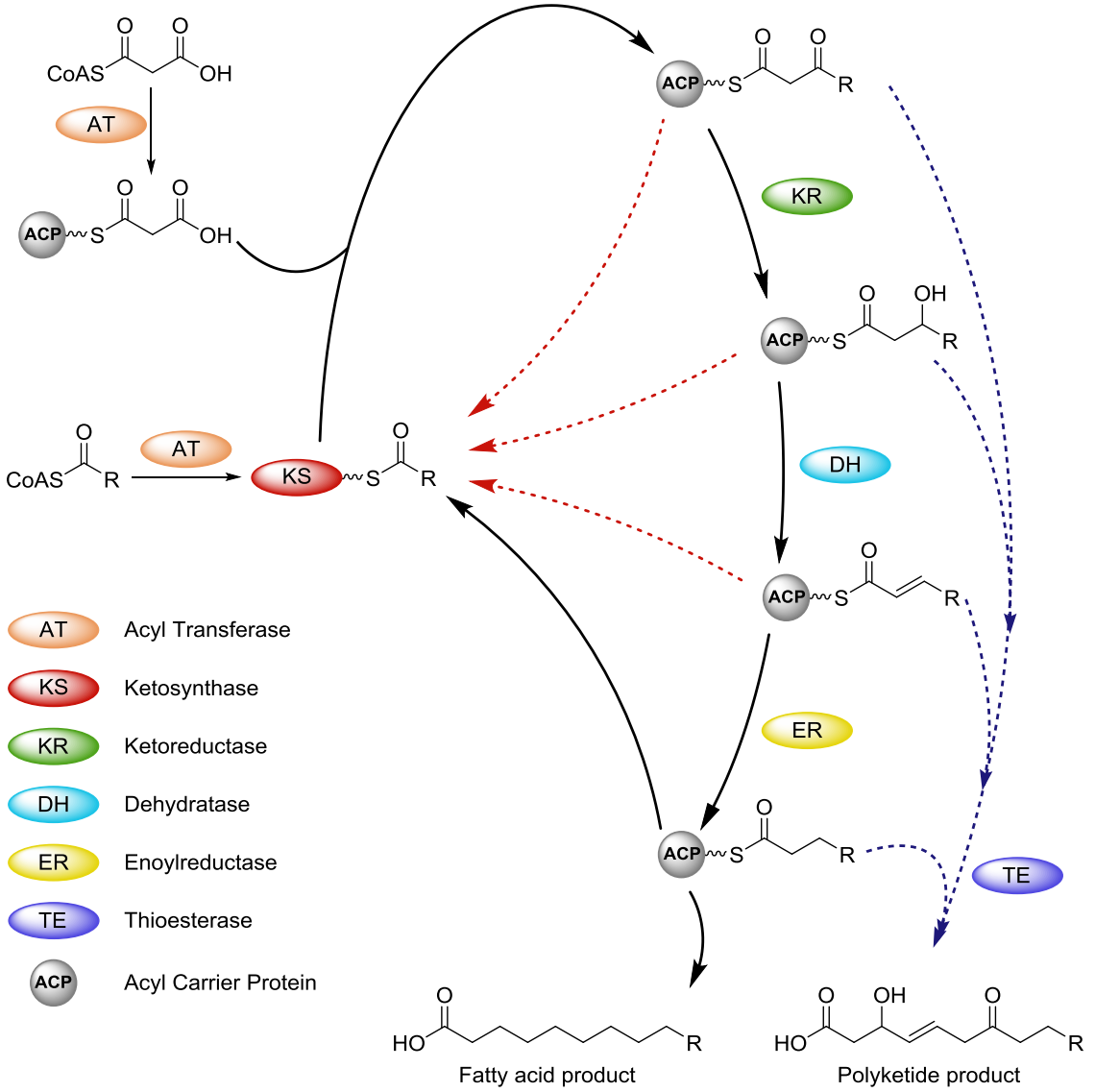
Scheme 2 ^{14}C isotopic labelling by Birch showing the incorporation of C_2 acetate units in aromatic polyketides.

Advances in genome sequencing, genetics and proteomic techniques have developed our understanding of biosynthetic gene clusters and the enzymology of the pathways they encode.[17–19] The diverse range of polyketide chemical structures arises from the broad range of enzymatic machinery and architecture available to synthesise these important molecules.[20,21]

1.3 Polyketide Biosynthesis

Polyketide biosynthesis proceeds from the transfer of an acyl-CoA substrate onto the thiol of an acyl carrier protein (ACP) as shown in Scheme 3.[17] Chain extension is carried out by a ketosynthase (KS) domain which acylates with the growing polyketide chain from an upstream ACP substrate onto a conserved cysteine residue. An acyl transferase (AT) domain transfers an extending unit (often malonyl CoA) onto the thiol of the downstream ACP, then decarboxylation and reaction with the KS-bound acyl group in a Claisen-like condensation results in carbon-carbon bond formation and a new β-keto moiety. The β-ketone may be reductively processed by a ketoreductase (KR) domain to give a β-hydroxyl, a dehydratase (DH) domain to give the α,β-unsaturated carbonyl and an enoylreductase (ER) to give the saturated alkyl chain. In an analogous catalytic cycle, fatty acid synthases (FAS) use all the reductive domains to produce saturated fatty acids, whilst polyketide synthases (PKS) may exhibit any number of processing domains and generate greater structural diversity.[22] Chain termination is achieved by a thiosesterase (TE) domain that may, in the presence of water, hydrolyse the polyketide to leave a terminal carboxylic acid or cyclise by macrolactonisation with a backbone hydroxyl group.

Further functionalisation of the linear backbone is achieved through alkylation at the α- or β-carbons. Substitution at the α-position is achieved by the incorporation of 2-substituted malonyl extender units or the action of *S*-adenosylmethionine (SAM) to methylate at the nucleophilic α-carbon to the thioester. Alkylation at the β-position is less common and is discussed in depth in Section 1.4.



Scheme 3 Fatty acid/polyketide biosynthesis. Fatty acids undergo full reductive processing of the β-ketone (solid black arrows) whereas polyketides exhibit a range of oxidation states according to domain architecture (dotted arrows). Intermediates can undergo chain extension without a full round of reductive processing (red) or may be released from the PKS by a thioesterase domain (blue).

PKSs may be categorised based on the structural arrangement of domains used for chain extension and processing (Figure 2).[23] Type I PKSs possess a series of catalytic domains covalently linked together to form a single polypeptide chain.[22] A set of domains that are required for a single round of chain extension and processing are referred to as a module. In type II PKSs, the catalytic domains exist as discrete proteins that associate into transient complexes to achieve polyketide biosynthesis.[24]

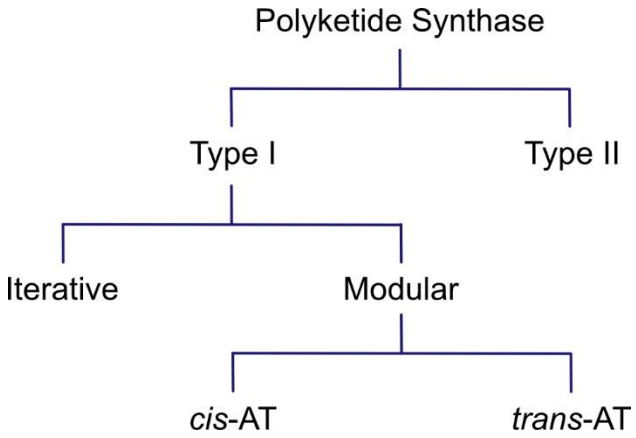
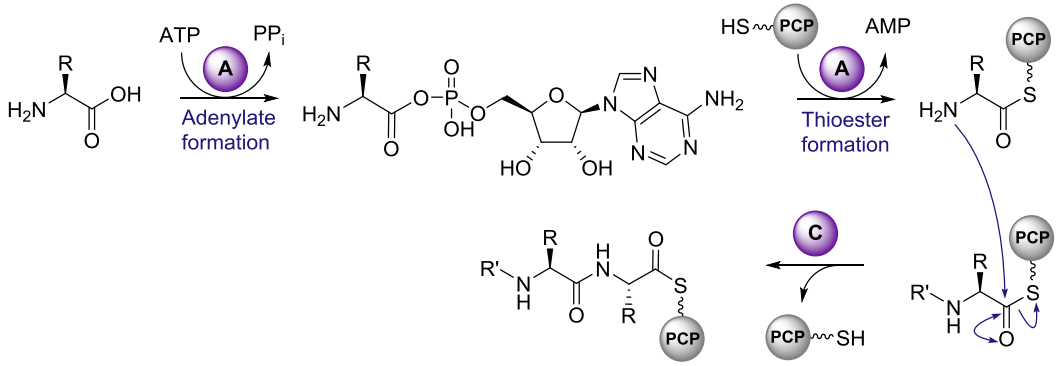


Figure 2 Schematic of the classification of polyketide synthases.

The type I PKS class may be further sub-divided into iterative and modular types.[25] Iterative PKSs are formed of a single repeat of catalytic domains that are utilised multiple times during biosynthesis.[26] Modular PKS are large, multi-functional polypeptides that contain all the catalytic domains required for complete elongation of the polyketide chain. Each domain is used only once as the substrate is extended and processed, passing downstream from the N-terminal to the C-terminal of the mega-enzyme.

The final sub-classification of type I modular PKSs depends on whether the AT domain is embedded in the modular system (*cis*-AT) or whether a free-standing AT (*trans*-AT) is required for the supply of extender units.[20,21,27] Inspection of domain order for a *cis*-AT PKS allows for the determination of the likely polyketide product, termed the co-linearity principle. However, this is not the case for *trans*-AT PKSs for which a wide-range of additional tailoring steps are often encountered.

Non-ribosomal peptide synthases (NRPS) exhibit many shared features with type I modular PKS. An extender unit (commonly an amino acid) is activated by an adenylation (A) domain and ATP to form a mixed anhydride (acyl-adenylate), which is then trapped by the thiol of a holo-peptidyl carrier protein (PCP) as shown in Scheme 4. Finally, a condensation (C) domain catalyses peptide bond formation between the extender unit and a PCP carrying the native chain. Due to the modular nature of PKS and NRPS, it is common to encounter products of hybrid PKS-NRPS pathways that contain both PKS and NRPS catalytic machinery.[28]



Scheme 4 Chain extension by a non-ribosomal peptide synthase (NRPS) involves an adenylation (A) domain to activate the amino acid extender unit and form a thioester bond with a peptidyl carrier protein (PCP). Amide bond formation is catalysed by a condensation (C) domain.

Cis-AT and *trans*-AT PKS differ not only in the architecture of the synthases, but they have evolved separately from the simpler FAS-like synthases.[29] The origins of *cis*-AT PKS arise from genetic replication of modules and subsequent modification of domain function. The *trans*-AT PKSs were genetically assembled by horizontal gene transfer between bacteria, resulting in a greater diversity in domain type and order.

This is exemplified by the 8 commonly found modular architectures for *cis*-AT PKS, whilst over 50 are known for *trans*-AT PKS.[27] Some of the peculiar features of *trans*-AT PKS include: unusual domain order, novel domain types, non-elongating modules, split modules, missing or silent domains and *trans*-acting tailoring domains. These features contribute to the remarkable diversity of the compounds isolated from the *trans*-AT PKS that are estimated to constitute 38% of bacterial modular PKS.[20,30]

Greater knowledge has been gained about the *cis*-AT PKS exemplified by the type I 6-deoxyerythronolide B synthase (DEBS) pathway that produces erythromycin, which has been extensively studied to understand the enzymology and structure.[21,31,32] More recently, back-to-back publications from Skiniotis *et al.* reported a series of cryo-EM structures for a module (PikAIII) from the *cis*-AT PKS that produces pikromycin.[33,34] The authors trapped multiple conformations during chain extension and reductive processing revealing the mobility of the ACP and conformational change required to achieve precise processing and directionality along the PKS.

In contrast, the *trans*-AT PKSs have been less well characterised structurally. In 2014, a solution-state SAXS structure along with homology modelling and computational analysis was undertaken on module 5 of VirA, a type I PKS that procures virginiamycin.[35]

This work was the first on an entire module from a *trans*-AT PKS and showed the overall conformation of the homo-dimeric module.

The small ACPs located within each module of the PKS are vital for correct fidelity of the assembly-line and must be post-translationally modified to convert the inactive apo-ACP into the activated holo-ACP. This requires the transfer of a phosphopantetheine (Ppant) arm onto a highly conserved serine residue at the base of helix II of the ACP (Figure 3). This 20 Å prosthetic arm bears a pendant thiol on to which growing acyl chain and extender building blocks are covalently attached *via* a thioester linkage. The ACP transfers the biosynthetic cargo to each catalytic domain for loading, chain extension and β-keto processing in a controlled and precise manner.

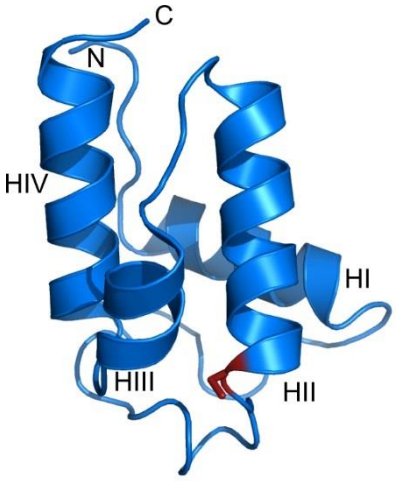


Figure 3 Actinorhodin ACP from *S. coelicolor* (pdb: 2k0y) with the conserved serine at the base of helix II in red.

1.4 β-Branching pathway

β-alkylation of the polyketide chain is less common than α-alkylation and requires attack of an alkyl nucleophile at the electrophilic β-position. Early isotopic feeding studies revealed C-2 of acetate as the origin for β-methyls in both mupirocin and virginiamycin.[36,37] Subsequent gene cluster analysis for polyketides bearing a β-methyl branch highlighted a conserved cassette of enzymes likely to facilitate the introduction of the alkyl branch. This cassette was termed a 3-hydroxy-3-methylglutaryl CoA-synthase (HCS) cassette, although this name may be used interchangeably with an HMGS cassette. Since these early reports, many compounds possessing β-branches have been reported and a selection of key β-branched compounds are shown in Figure 4.

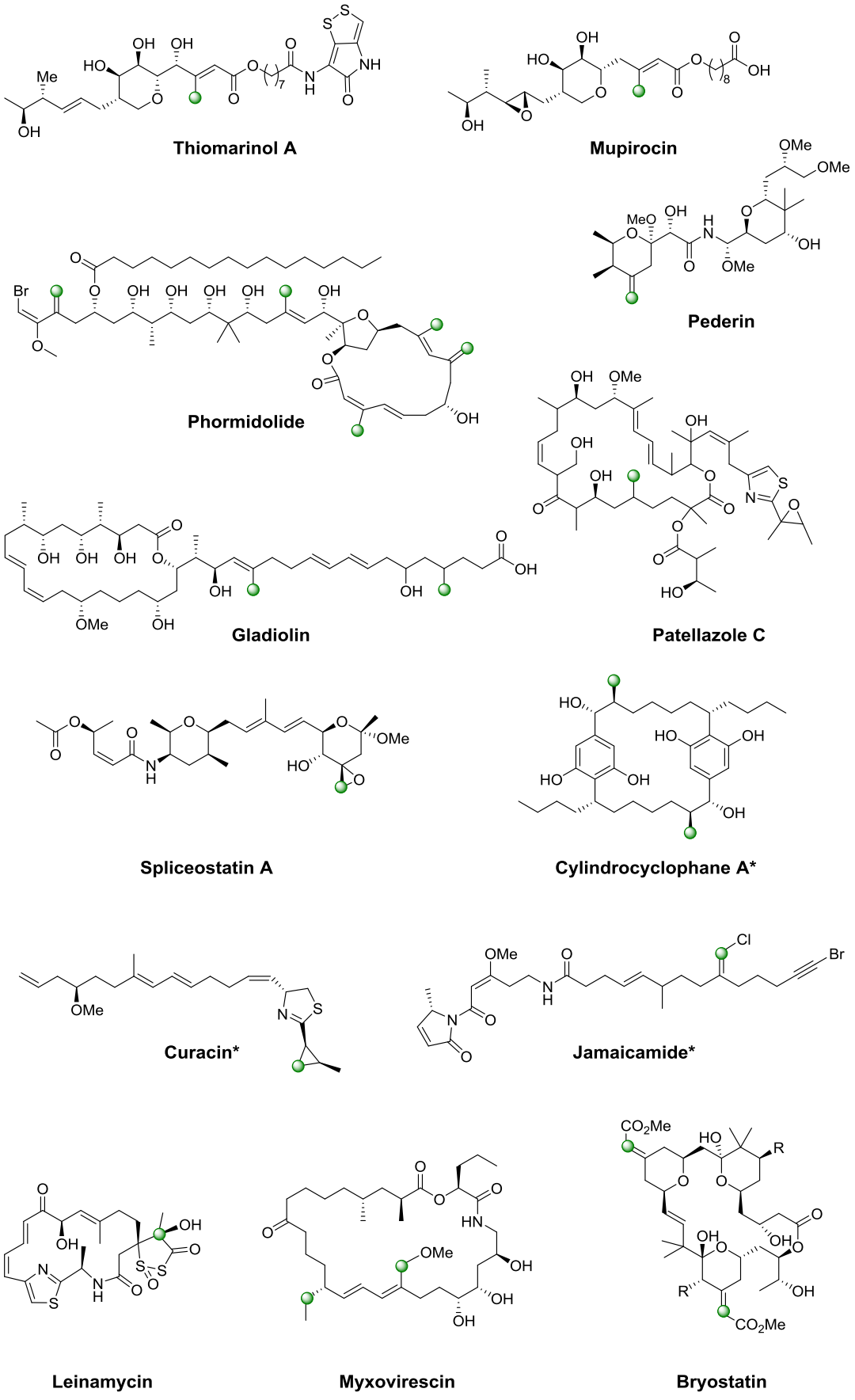
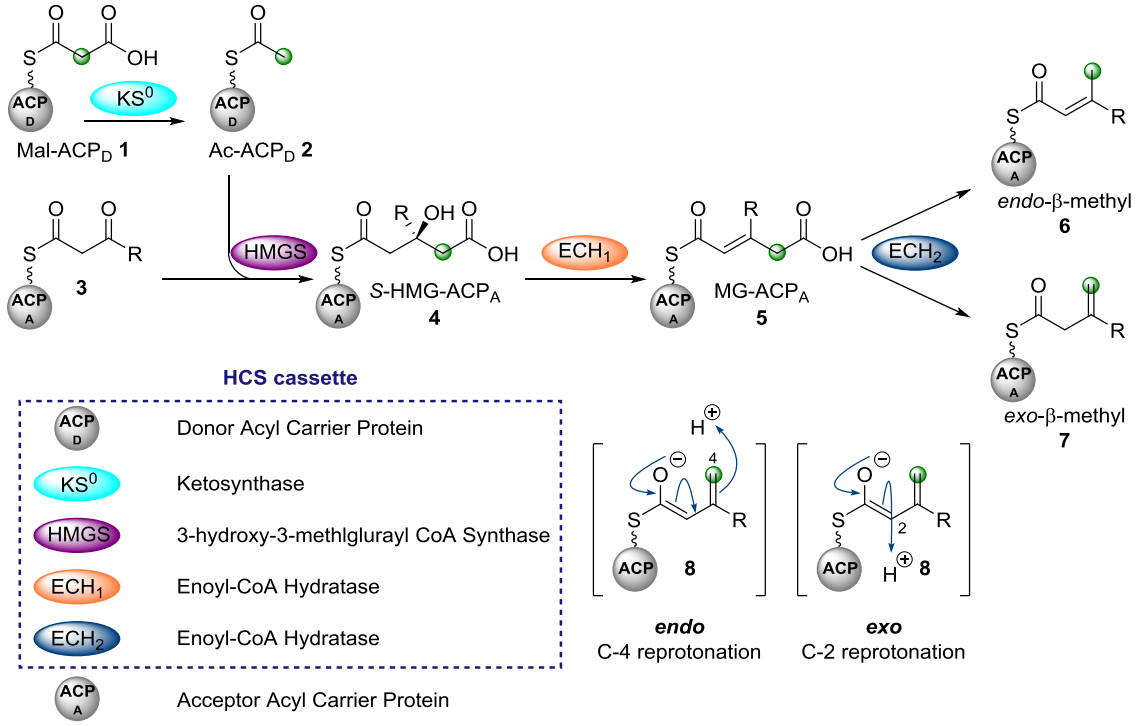


Figure 4 Examples of polyketides containing β-branches from *cis*-AT (*) and *trans*-AT PKS. The green dot shows the position of incorporation of acetate C-2 introduced by the HCS cassette.

β-branching is most commonly observed in *trans*-AT PKS pathways, although several notable *cis*-AT polyketides exhibit β-branches, such as curacin and jamaicamide.[38] In both systems, they arise through the interaction of an HCS cassette with a β-ketothioester biosynthetic intermediate bound to a modular acceptor ACP (ACP_A). The minimal cassette contains a free-standing donor ACP (ACP_D), a KS lacking the conserved cysteine residue required for condensation (KS^0) and a 3-hydroxy-3-methylglutaryl synthase (HMGS). Two enoyl CoA hydratase (ECH) domains that belong to the crotonase superfamily (CS) of enzymes complete the standard set of five enzymes. However, additional tailoring domains may also be present and influence the chemical moiety introduced by the cassette.

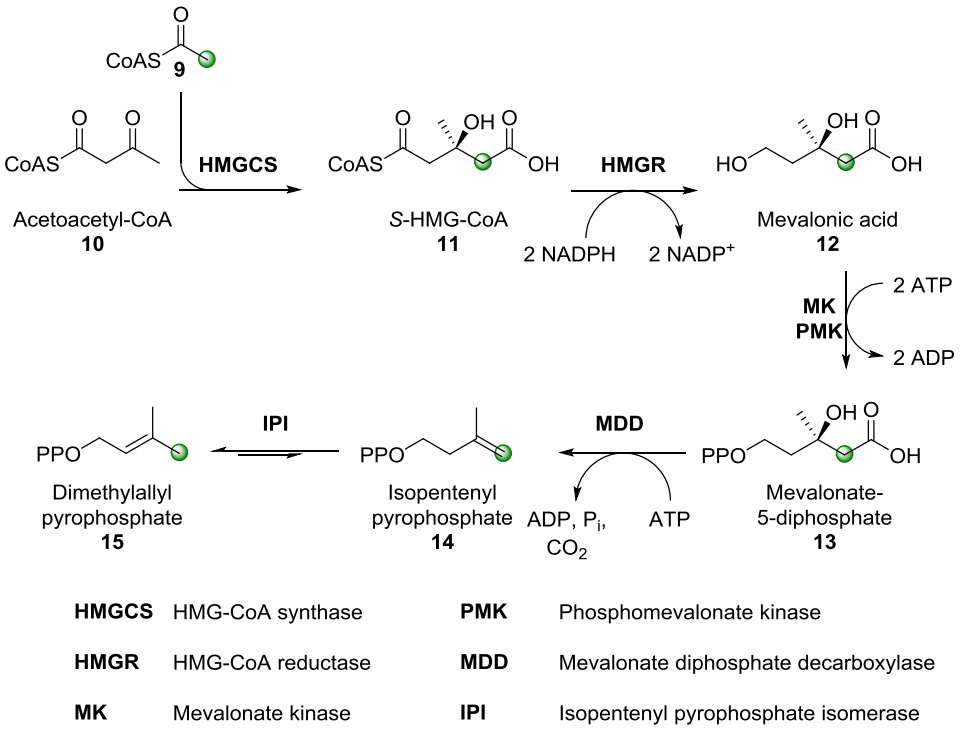


Scheme 5 Mechanism for the incorporation of an *endo*-β-methyl **6** or *exo*-β-methyl **7** branch by an HCS cassette. Colour coding is used to indicate domain type and is used consistently throughout this thesis.

The mechanism of β-branching in the bacillaene biosynthetic pathway was elucidated by Calderone and co-workers using biochemical assays and mass spectrometry.[39] Malonyl CoA loads the free-standing ACP_D to give Mal-ACP_D **1** which is then decarboxylated by KS^0 to give Ac-ACP_D **2** (Scheme 5). The acetyl group is then transferred to the active site cysteine of the HMGS, deprotonated at the α-carbon, and the resulting enolate attacks the β-ketone of a β-ketothioester **3** bound to ACP_A in an aldol reaction. The HMGS is subsequently hydrolysed to give HMG-ACP_A **4**, the key intermediate in the β-branching

pathway. Consistent with the formation of *S*-HMG-CoA in mevalonate-dependant isoprene biosynthesis, *S*-HMG-ACP_A is the product of the condensation.[40–42] Dehydration by ECH_1 gives MG-ACP_A **5** and decarboxylation by ECH_2 furnishes an unsaturated β-branched **6** polyketide.

The most common isomer produced by an HCS cassette is an α,β-unsaturated β-branch (*endo*-β-methyl, **6**) with the tri-substituted *endo* double bond arising from the reprotonation of the C-4 carbon of **8** by ECH_2. The alkene product is often characterised by UV spectroscopy for the α,β-unsaturated thioester chromophore.[39,43] Decarboxylation and reprotonation at C-2 of **8** by an ECH_2 domain gives rise to the β,γ-unsaturated β-branch (*exo*-β-methyl, **7**) and has been characterised in the formation of the vinylchloride moiety in jamaicamide.[44,45] The *endo*-β-methyl and *exo*-β-methyl nomenclature is used throughout this thesis to indicate the position of unsaturation in β-branched polyketides.

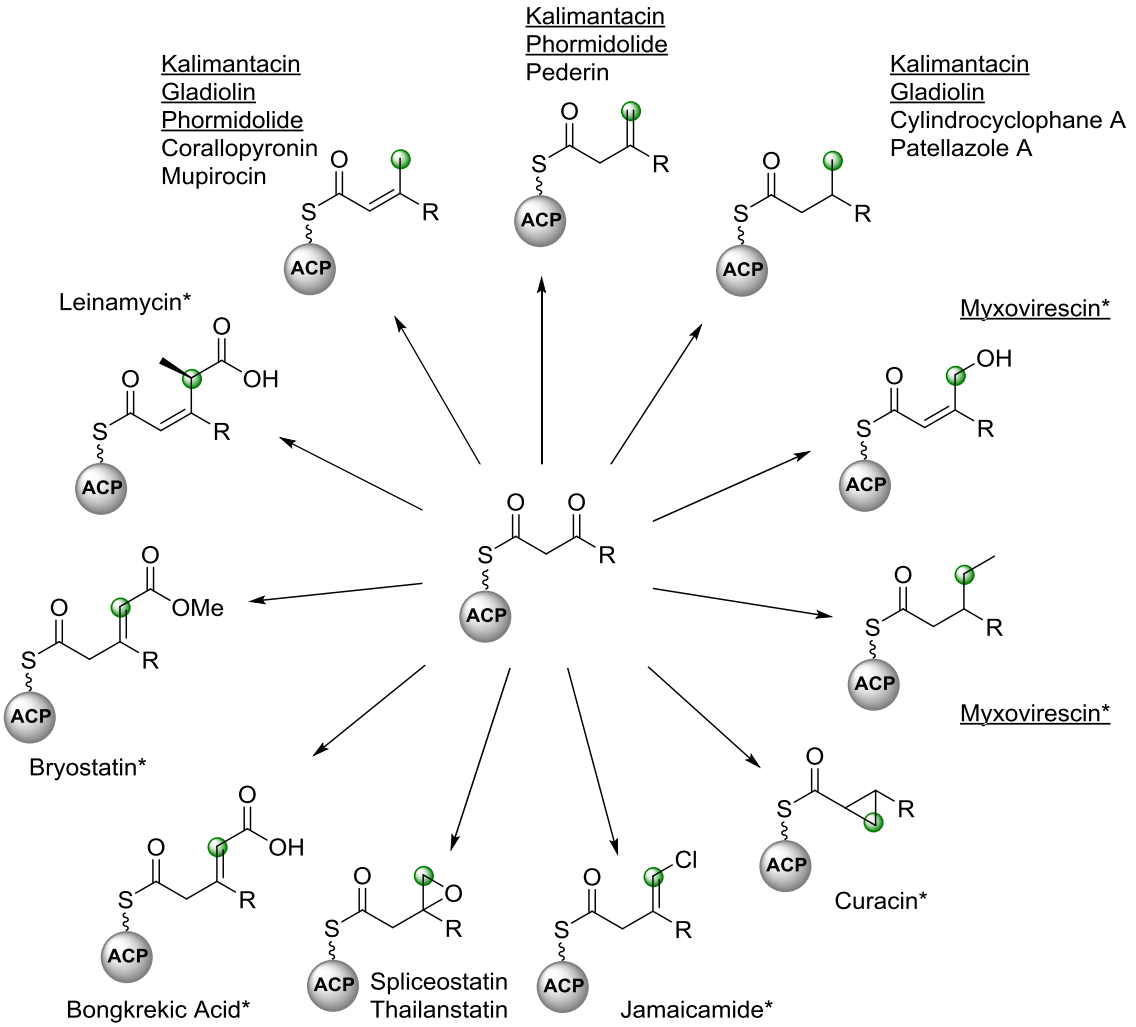


Scheme 6 Mechanism of mevalonate-dependant isoprene biosynthesis, a pathway with shared mechanistic features to polyketide β-branching.

Comparisons may be drawn between polyketide β-branching and mevalonate-dependent isoprenoid biosynthesis from primary metabolism (Scheme 6).[38,42,46] In this pathway Ac-CoA **9** and Acac-CoA **10** form *S*-HMG-CoA **11** in an aldol condensation catalysed by HMG-CoA synthase (HMGCS), a homolog of HMGS. HMG-CoA is then reduced by HMG reductase (HMGR), cleaving the CoA- thioester to give mevalonic acid **12**. The

primary alcohol is doubly phosphorylated by mevalonate kinase (MK) and phosphomevalonate kinase (PMK) to give **13**. Mevalonate diphosphate decarboxylase (MDD) activates the tertiary alcohol prior to a decarboxylative elimination to produce isopentyl pyrophosphate **14**. Isopentyl pyrophosphate isomerase (IPI) catalyses the isomerism to dimethylallyl pyrophosphate **15**. These two simple 5-carbon building blocks are utilised in the biosynthesis of terpenes.[47,48]

A wide range of alternative processing has been identified and characterised for polyketide β-branching and a summary of all known β-branches, as of January 2019, is shown in Scheme 7. The most common pattern of branching is the introduction of a single β-branch (mupirocin) or two identical β-branches (corallopyronin). However, there are examples of compounds that contain different β-branches, which require PKS selectivity to ensure fidelity of β-branching, for example gladiolin, phormidolide and kalimantacin.



Scheme 7 All β-methyl branches reported in the literature and examples of natural products that incorporate each branch. Those with a star (*) are the only example whilst those underlined possess more than one type of β-branch.

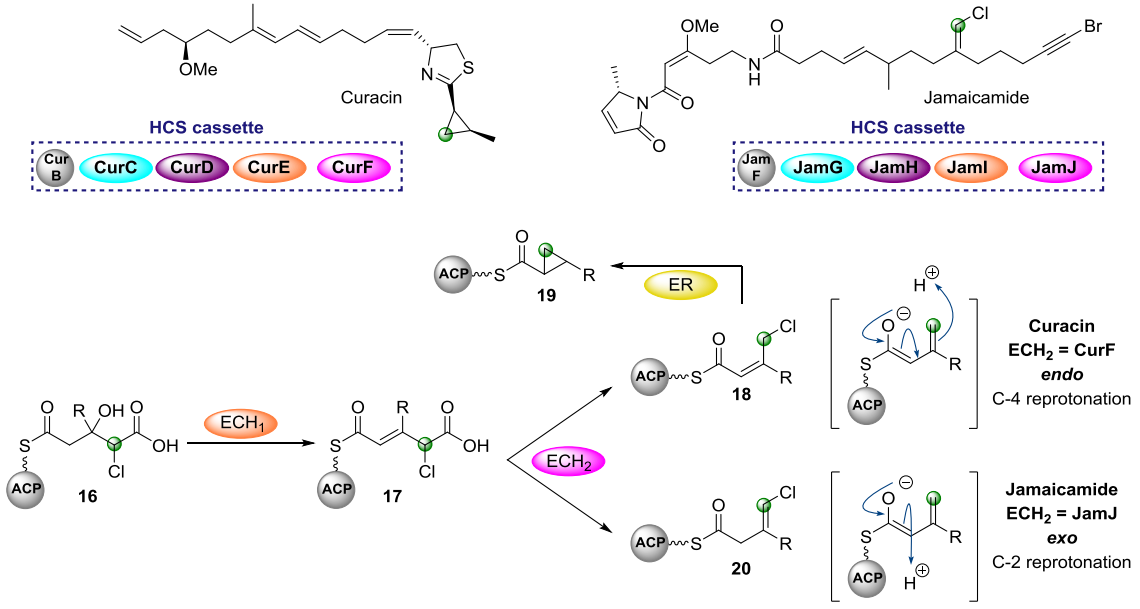
The role of ACPs during β-branching

The mechanism of β-branching is well characterised and the diversity of β-branching architecture demonstrates the control required to prevent aberrant modifications. The key to achieving selectivity is through the delivery of the correct substrate to each catalytic domain by the ACP. ACP_A is required to interact with the adjacent KS domains of the modular PKS, as well as the *trans*-acting AT, HCS cassette (HMGS/ECH_1/ECH_2) and additional tailoring enzymes. ACP_D on the other hand, must only interact with the HMGS, placing a stricter requirement for recognition. Phylogenetic analysis shows that ACP_A and ACP_D clade separately from each other.[49] Through co-crystallisation of ACP_D (CurB) and HMGS (CurD) from the curacin pathway, the charge and shape complementarity of ACP_D arising from an unusual helix III position has been shown.[41] The affinity of HMGS for ACP_D was higher than ACP_A and coupled with the presence of two ACP_D/HMGS pairs that install different β-branches in the biosynthesis of myxovirescin,[50] the donation of the correct substrate by ACP_D is key to achieving precise β-branching.

ACP_A may be distinguished from other PKS ACPs by a conserved tryptophan positioned six residues after the serine for Ppant attachment (GxDSxxxxxW).[51] The burial of this tryptophan creates a hydrophobic core that distinctively presents helix II and III for interaction with the HMGS. Non-branching PKS modules normally contain a single ACP and cryo-EM structures of a whole module from pikromycin biosynthesis elegantly showed the architecture and movement of the ACP to the modular catalytic domains.[33,34] However, β-branching modules may contain between 1 to 3 copies of ACP_A. Tandem ACPs in the mupirocin gene cluster have been shown *in vivo* to increase the flux through the pathway at a rate-limiting, *in-trans* biosynthetic step.[52] This view was reinforced by an NMR structure of the di-domain ACP_A from mupirocin that showed two, independent domains that were predicted to function in parallel.[51] Studies on the curacin biosynthetic pathway demonstrated *in vitro* an increased efficiency with tandem ACPs, however, some synergistic interaction of the triplet domain was also observed.[53] Recently, Kim *et al.* used genetic engineering of the DEBS modular PKS to install a module containing a triplet ACP compared to a single ACP in the wild-type. A 2.5x increase in production was observed, further suggesting that increasing ACP concentration results in higher flux and polyketide production by a modular PKS.[54]

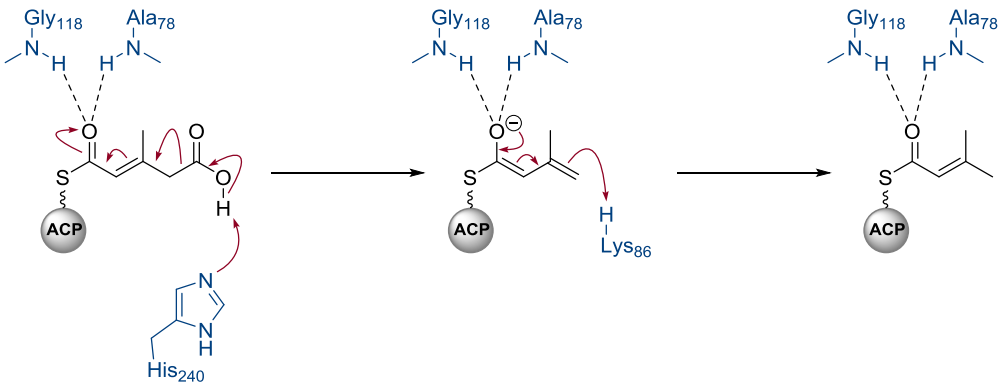
Alternative β-branching

Canonical β-branching uses a full 5-enzyme cassette to produce *endo*-β-methyl branches, however, a range of β-branches have been reported. Their biosynthesis may rely on modified (ECH_2), absent (ECH_2) or additional (ER) domains. Smith *et al.* have undertaken extensive work on the HCS cassettes from the related *cis*-AT polyketides curacin and jamaicamide (Scheme 8). They initially reported the coupling of the ECH_1/ECH_2 catalysed steps from the curacin pathway to furnish an *endo*-β-methyl branch.[40] The crystal structure of CurF ECH_2 was then reported along with a proposed mechanism for the formation of the *endo*-β-methyl branch (Scheme 9).[55] The preference of CurF ECH_2 for ACP-bound intermediates over CoA-bound intermediates was also highlighted.



Scheme 8 β-branching pathway in the related *cis*-AT PKS compounds curacin and jamaicamide. Following ECH_1-catalysed dehydration, the pathways diverge to form a chlorinated *endo*-β-methyl (curacin) that is subsequently cyclised to a cyclopropane ring, and an *exo*-β-methyl (jamaicamide) catalysed by ECH_2 domains CurF and JamJ respectively. The pink ECH_2 annotation refers to a modular configuration of the domain.

A key publication in 2009 highlighted the divergence between the related biosynthetic pathways of curacin and jamaicamide.[45] In curacin, following formation of an *endo*-β-methyl branch, a modified ER downstream of ECH_2 catalyses cyclopropane formation of chlorinated intermediate **18** to give **19** (Scheme 8). In jamaicamide, the product of ECH_2 decarboxylation is *exo*-vinyl chloride **20** that results from C-2 reprotonation. The identity of the two β-branches was confirmed by hydrolysis of ACP-bound products as amides, GC/MS analysis and comparison to synthetic standards.



Scheme 9 Mechanistic proposal for the formation of an *endo*-β-methyl by CurF ECH_2. His_{240} deprotonates the carboxylic acid, followed by the loss of CO_2 and reprotonation by Lys86. Ala78 and Gly118 stabilise the enolate in an oxyanion hole.

Pederin and a series of structurally related compounds that each contain an *exo*-β-methyl branch have been reported.[44,56–59] The *exo*-β-methyl branch is installed by an HCS cassette in which the only ECH_2 domain is immobilised adjacent to the ACP_A as part of the modular PKS (Figure 5A). Although not proven, the likely mechanism for the formation of the *exo*-β-methyl is through ECH_2-catalysed decarboxylation and subsequent C-2 reprotonation akin to jamaicamide biosynthesis. Spliceostatin and thailanstatin are additional compounds that from an *exo*-β-methyl branch with a single, modular ECH_2 domain, which undergo subsequent oxidation to install an epoxide moiety (Figure 5B).[60,61] *Exo*-β-methyl branch installation is discussed in more detail in Section 2.1.1.

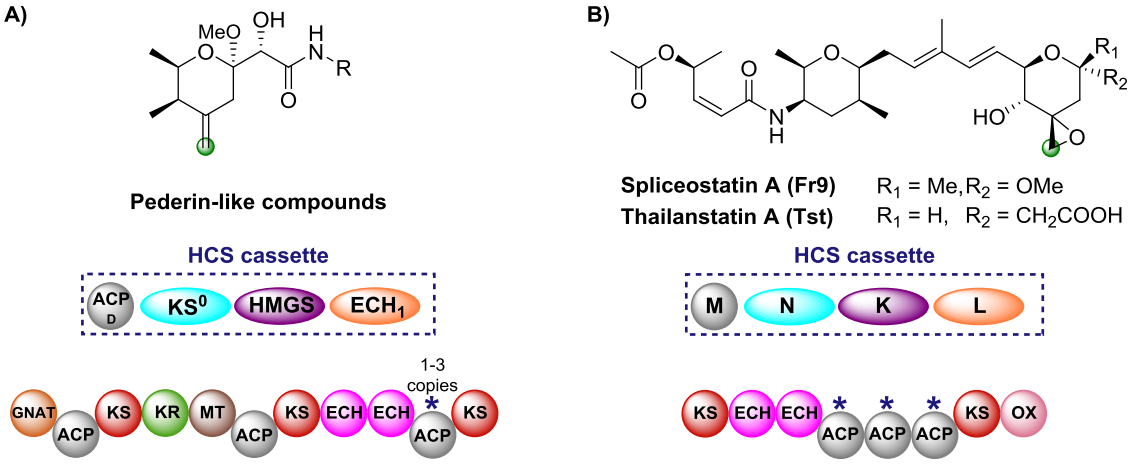


Figure 5 The formation of *exo*-β-methyl branches by single, modular ECH_2 domains. β-branching ACP_A are labelled (*). A) The shared modular architecture of pederin and the related compounds onnamide, diaphorin, nosperin and psymberin. B) Epoxidation of an *exo*-β-methyl by a flavin-dependent monooxygenase (Ox) results in epoxide moieties in spliceostatin and thailanstatin.

Installation of an *endo*-β-methyl branch generates an α,β-unsaturated thioester that may be reduced to the saturated alkyl chain by an ER. In patellazole C biosynthesis, a saturated-β-methyl is incorporated by an unusual HCS cassette lacking a *trans*-acting ADP_D (Figure 6).[62] The modular PKS PtzD contains a di-domain β-branching ACP_A but also a candidate for ACP_D located at the C-terminus of the polypeptide. This ACP showed homology to PedN, the free-standing ACP_D from the pederin gene cluster, providing a rare example of an in-*cis* ACP_D. PtzQ is a trans-acting ER domain that is hypothesised to carry out the reduction of the *endo*-β-methyl to produce the sat.-β-methyl branch.

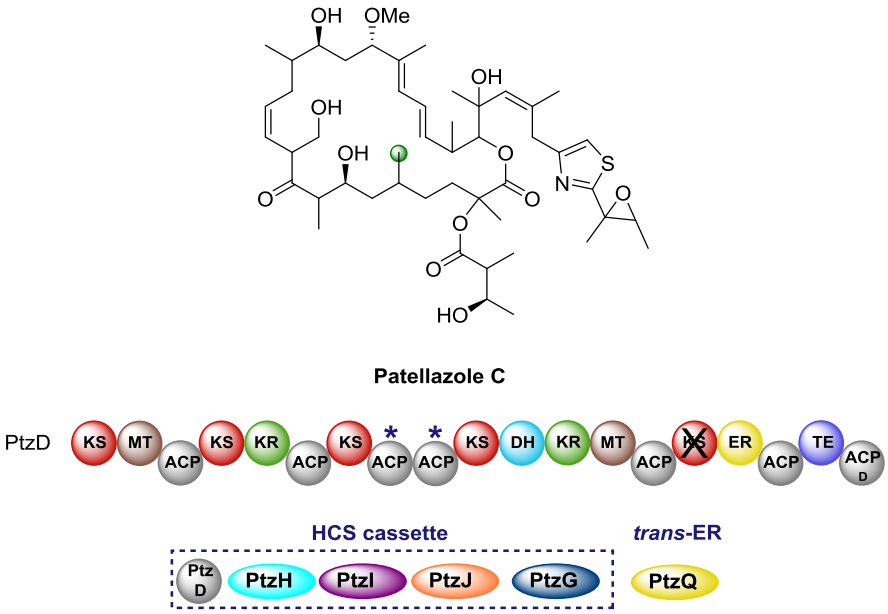
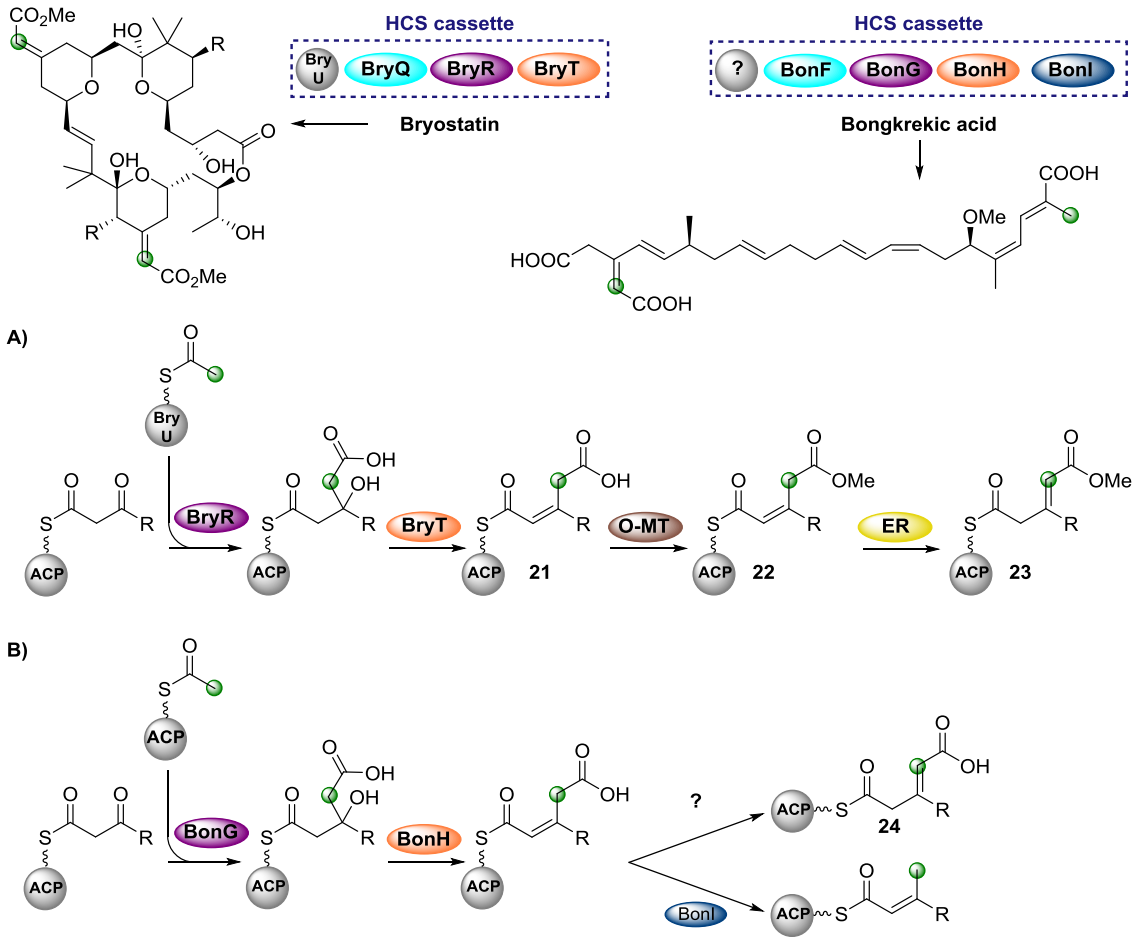


Figure 6 Patellazole C contains a sat.-β-methyl branch. PtzGHIJ form the *trans*-acting HCS cassette, however, ACP_D was not identified as a free-standing domain. A candidate, with homology to PedN from the pederin gene cluster, was identified at the C-terminus of PtzD. The *trans*-acting ER PtzQ is hypothesised to perform the reduction. β-branching ACP_A are labelled (*).

Whilst virtually all β-branching pathways involve an ECH_2 decarboxylation, bryostatin and bongkrekic acid are two polyketides that lack this enzymatic step. A recent publication showed that in bryostatin biosynthesis, the α,β-unsaturated (relative to the thioester) intermediate **21** was produced as the result of ECH_1-catalysed dehydration which was then specifically *O*-methylated to give **22** (Scheme 10A).[63] It is hypothesised that a downstream ER may catalyse the isomerism to **23** prior to tetrahydropyran ring formation by an unknown mechanism. Similarities in β-branching structure have been drawn between bryostatin and bongkrekic acid, which contains two β-branches; an *endo*-β-methyl and a non-decarboxylated acrylate **24** as confirmed by isotopic feeding studies (Scheme 10B).[64] A full HCS-cassette was found in the gene cluster with no obvious

mechanism for the selective installation of each β-branch.[65] Bongkrekic acid is an interesting example of divergence of pathways from a common intermediate to introduce the two different β-branches within one biosynthetic assembly line. The lack of off-loading domain from the modular PKS led to a currently unproven and unique hypothesis of concurrent β-branching and off-loading.



Scheme 10 Bongkrekic acid and bryostatin contain non-decarboxylated β-branches. A) Generation of *O*-methylated β,γ-unsaturated β-branch (**23**) by ECH_1-catalysed α,β-dehydration, *O*-methylation and isomerism by a modified ER domain. B) Differential formation of an *endo*-β-methyl and non-decarboxylated β-branch (**24**) by the HCS cassette in bongkrekic acid biosynthesis is less well understood. The identity of ACP_D has not been reported.

In some cases, multiple, different β-branches may be incorporated into the same compound requiring the HCS cassette and tailoring enzymes to selectively install each β-branch with high fidelity. Thus, the HCS cassette must only interact with tryptophan-flagged β-branching ACP_A and the correct branch must be introduced at each ACP_A. The related compounds gladiolin and etnangien share many structural features but differ in their β-methyl incorporation pattern (Figure 7).

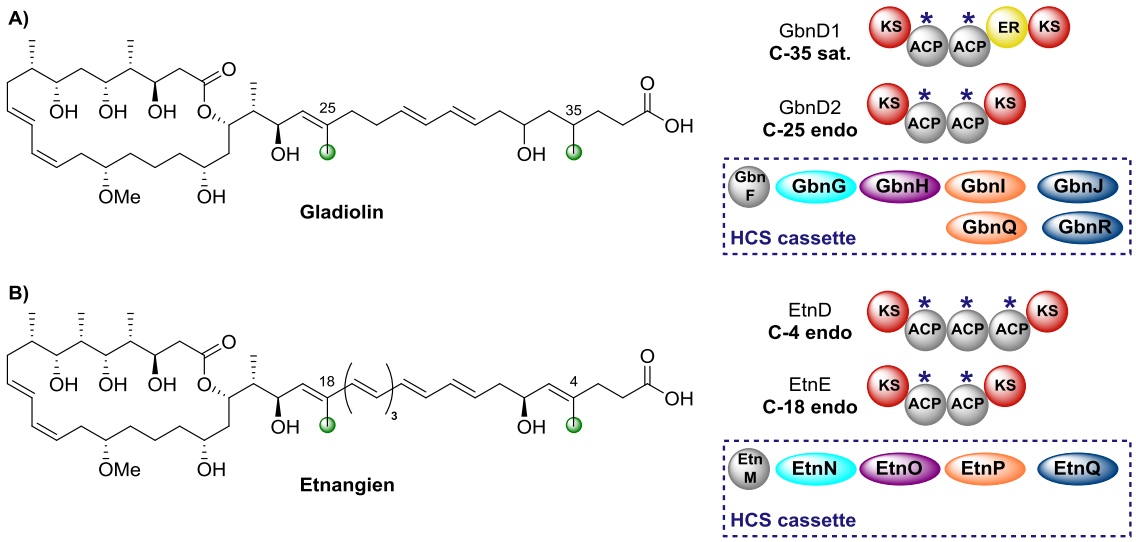


Figure 7 A) Gladiolin contains an *endo*-β-methyl (C-25) and sat.-β-methyl branch (C-35). The *trans*-acting HCS cassette contains two pairs of ECH_1/ECH_2 domains and an in-*cis* ER located in module 1 of GbnD1 to reduce the α,β-unsaturated β-branch. B) Etnangien contains two *endo*-β-methyl branches (C-4 and C-18) that are installed by a single, HCS cassette. β-branching ACP_A are labelled (*).

Etnangien contains two *endo*-β-methyl branches (C-4 and C-18) incorporated by a 5-enzyme HCS cassette.[66] Other than selectively interacting with the β-branching ACPs, the cassette shows no discrimination between the two sites of β-methyl incorporation. In gladiolin biosynthesis, the gene cluster encodes a *trans*-acting HCS cassette, that contains two sets of ECH_1/ECH_2 domains, to incorporate an *endo*-β-methyl (C-25) and sat.-β-methyl branch (C-35).[67] An ER in module 1 of GbnD1 is located at the point of sat.-β-methyl incorporation to carry out the reduction. The presence of two sets of ECH_1/ECH_2 domains led the authors to speculate on whether the modular ER domain places structural constraints on the β-branching ACP_As requiring the recruitment of a second set of ECH_1/ECH_2 domains.

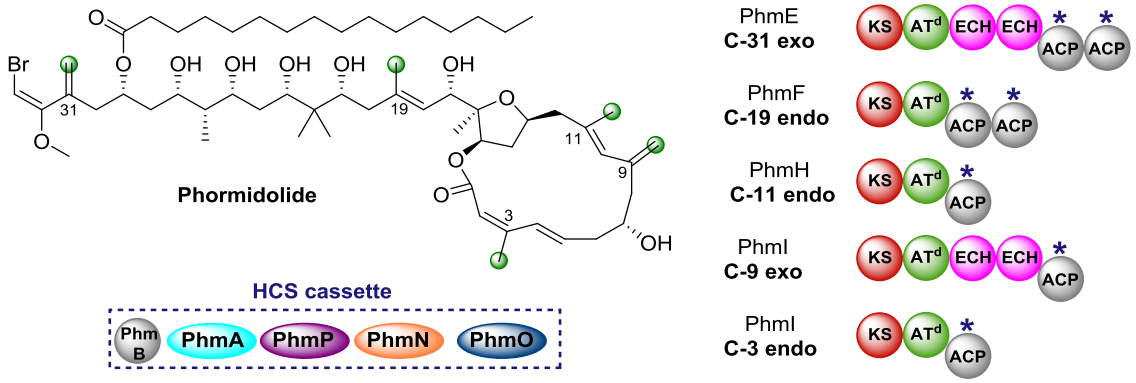
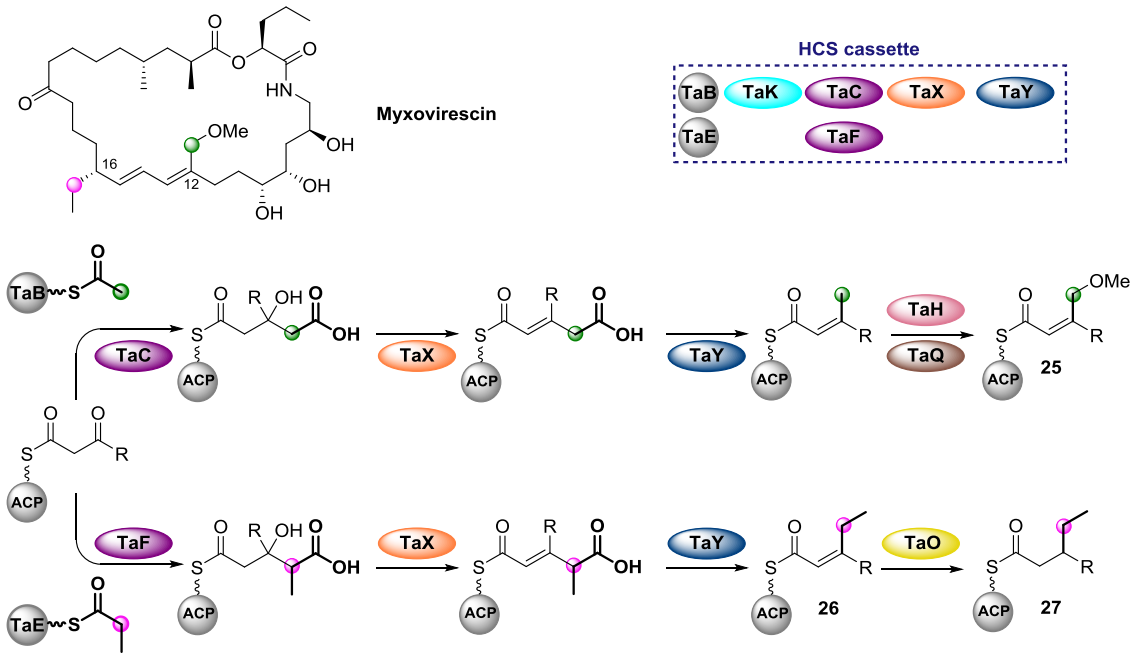


Figure 8 Phormidolide contains five β-methyl branches introduced by one *trans*-acting HCS cassette and tailoring ECH domains located in PhmE and PhmI at the point of *exo*-β-methyl incorporation. β-branching ACP_A are labelled (*).

Phormidolide contains the greatest number of β-branches in a single compound currently reported in the literature, comprising of three *endo*-β-methyl and two *exo*-β-methyl branches (Figure 8).[68] An HCS cassette, as well as modular ECH domains adjacent to the β-branching ACPs, are thought to introduce the different types of β-branch. However, the authors state that the role of the modular ECH domains remained unknown.

One of the most complex β-branching cassettes is found in the gene cluster of myxovirescin (Scheme 11).[38] The natural product contains two different β-branches: an *endo*-β-methoxymethyl **25** at C-12 arising from hydroxylation (TaH) and *O*-methylation (TaQ) of an *endo*-β-methyl, and an unusual sat.-β-ethyl branch **27** at C-16 from reduction (TaO) of an *endo*-β-ethyl branch **26**. There are two pairs of ACP_D/HMGS domains in the gene cluster, however, only single copies of KS^0, ECH_1 and ECH_2. *In vivo* feeding studies with labelled precursors established the origin of the C-12 branch to be acetate and the C-16 branch to be methylmalonyl/propionate.[69] *In vitro* reconstitution of the HCS cassette demonstrated two translationally coupled ACP_D/HMGS pairs whereby TaB/TaC selectively introduce the C-12 branch and TaE/TaF the C-16 branch,[43] which was supported by *in vivo* gene deletions.[70]

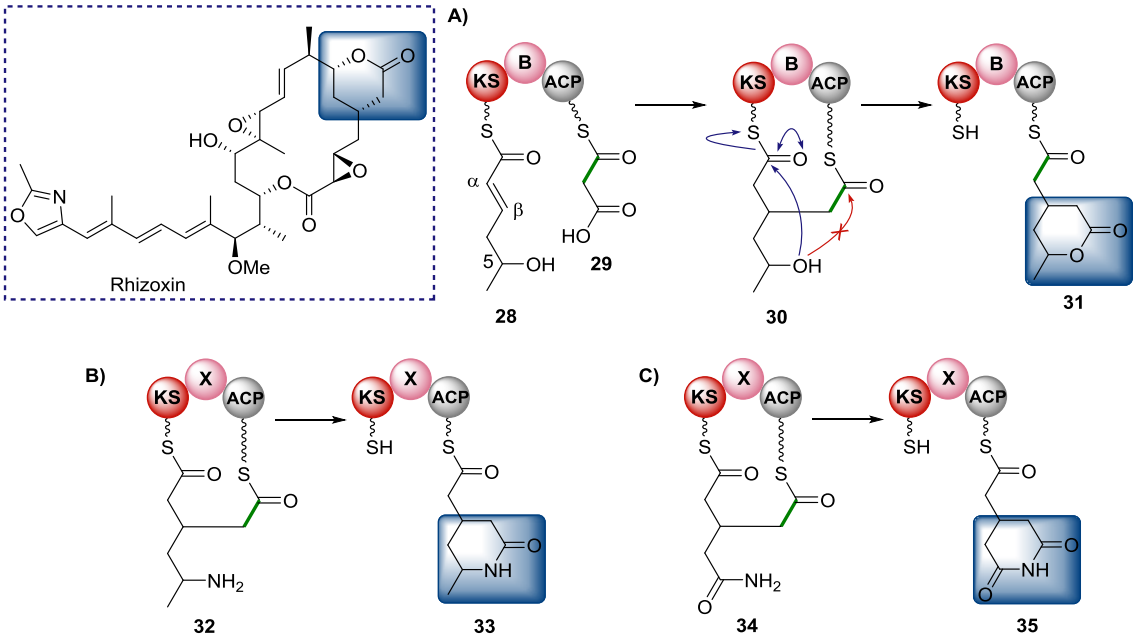


Scheme 11 Incorporation of methyl (green) and ethyl (pink) *endo*-β-branches in the biosynthesis of myxovirescin. The gene cluster encodes two ACP/HMGS pairs that select and react either acetyl-ACP (TaB/TaC) or propionyl-ACP (TaE/TaF) with the modular β-ketothioester. Dehydration and decarboxylation are carried out by the same ECH_1/ECH_2 pair (TaX/TaY). TaH (oxidation) and TaQ (*O*-methylation) produce the *endo*-β-methoxymethyl at C-12 **25** and TaO (ER) reduces the α,β-alkene to produce the sat.-β-ethyl branch **27** at C-16.

Furthermore, a Δ*TaF* mutant produced a compound with *endo*-β-methyl branches at C-12 and C-16, showing TaF can complement for TaC.[50] However, the opposite complementation is not observed, and these results were further supported *in vitro* whereby TaC may react with both ACP_As, whilst TaF only reacts with its cognate ACP_A. Overall, two distinct sets of enzymatic machinery are utilised in the formation of different β-branches, with the shared steps occurring only once the HMGS-catalysed condensation has occurred.

β-alkylation by Michael addition

β-alkylation of polyketides may also be achieved through a mechanistically different pathway.[71] In the biosynthesis of rhizoxin, a two-carbon unit is introduced at the β-position that then cyclises to form a δ-lactone (Scheme 12).[72] An unprecedented modular architecture (KS-B-ACP) was observed with a newly defined branching domain (B). *In vitro* reconstitution of the module, X-ray crystallography, protein cross-linking and stable isotope studies were used to elucidate the mechanism of branching. An ACP-bound malonyl group **29** is decarboxylated and attacks the β-position of a KS-bound α,β-unsaturated thioester **28**. The C-5 hydroxyl of **30** subsequently attacks the KS-thioester to liberate the δ-lactone product **31**.

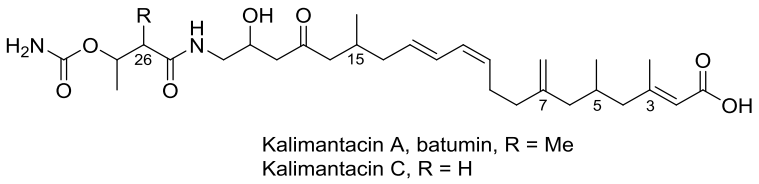


Scheme 12 Bacterial polyketide rhizoxin contains an unusual two-carbon β-branch. A) The mechanism for the formation of the δ-lactone by Michael addition onto the unsaturated α,β-thioester followed by cyclisation. B) The amino variant in the synthesis of lactams. C) The formation of pharmacologically-important glutarimide moiety by attack of a carboxamide.

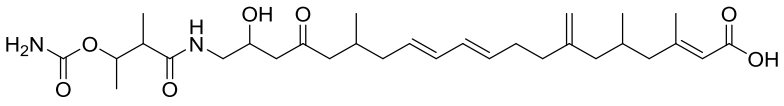
Investigation into the scope of the reaction revealed amine **32** and carboxamide **34** nucleophiles could be used to synthesise lactam **33** and glutarimide analogues **35**.[73] Genetic analysis of the glutarimide-containing polyketides revealed a homolog to the B-domain, that was labelled an X-domain. A range of chimeras were constructed and used to demonstrate complementation between B/X-domains and the structural rather than catalytic role they play in the formation of β-branched polyketides.[74]

1.5 Kalimantacin

The kalimantacins were first isolated from *Alcaligenes* sp. YL-02632S by Kamigiri and co-workers in 1996 (Figure 9).[75] The major product isolated was kalimantacin A which contains an *E,Z*-diene and C-26 methyl group. Minor compounds kalimantacin B containing an *E,E*-diene and kalimantacin C lacking the C-26 methyl were also isolated. The bio-activity of these compounds was reported and kalimantacin A was shown to have antibiotic activity against *S. aureus* and *S. epidermis*.[76]



Kalimantacin A, batumin, R = Me
Kalimantacin C, R = H



Kalimantacin B

Figure 9 Structure of kalimantacin A-C originally isolated from *Alcaligenes* sp. YL-02632S. Batumin from *Pseudomonas batumici* and kalimantacin from *Pseudomonas fluorescens* strain BCCM_ID9359 have subsequently been shown to be kalimantacin A.

In 2000, Smirnov *et al.* reported the isolation of batumin from *Pseudomonas batumici*, harvested from the Black Sea, that was reported to have the same molecular structure as kalimantacin A.[77] No further reports were made on these or similar compounds until 2010 when Lavigne *et al.* reported the isolation and characterisation of a compound from *Pseudomonas fluorescens* strain BCCM_ID9359.[78] The structure elucidated by HRMS and NMR corresponded to the structure of kalimantacin A and batumin. Due to the presence of 5 unknown stereocenters it was not clear whether kalimantacin A and batumin were the same compound. Each of the compounds isolated showed strong, selective anti-bacterial properties against staphylococci species, including multi-drug resistant strains,[75] and to date no resistant strains have been reported.[78,79]

Lavigne *et al.* identified the gene cluster responsible for the production of kalimantacin from *P. fluorescens* through a combination of BLAST searches and gene knock-out experiments.[78] Sequence analysis identified three large open reading frames (ORF) and a series of smaller ORFs encoding a type I PKS/NRPS hybrid synthase and tailoring enzymes. From the module and domain analysis it was determined that ACPs were loaded by free-standing AT domains, thus falling into the *trans*-AT PKS classification. A subsequent comparison of the gene clusters from *P. fluorescens* and *P. batumici* revealed they were 100% identical and that both kalimantacin A and batumin are the same compound. The name kalimantacin is used throughout this thesis.[80]

The biosynthetic pathway presented in Figure 10 includes several updates compared to the original publication.[78] Firstly, BatH and BatJ were both labelled as AT domains with the proposal that BatH transfers the starting acetyl unit and BatJ loads malonyl extender units. However, recent work has shown that BatH is an acyl hydrolase (AH) thought to be involved in proof-reading.[81] Whilst AH domains are related to AT domains, they clade separately and their hydrolytic activity has been determined *in vitro*. Secondly, Keatinge-Clay *et al.* proposed a paradigm-shift in the annotation of modules within *trans*-AT PKS.[82] Classically, the ACP is used to define the C-terminal domain of each module, with the KS and tailoring domains upstream of the ACP. However, recent work has established that *trans*-AT PKS evolved from horizontal gene transfer and that there is a co-evolution of the ACP with the downstream KS domain.[29,83] Keatinge-Clay presented a comprehensive analysis of modular ACPs and demonstrated that those from a common modular architecture clade together. As a result, it has been proposed that the ACP domain should sit central within a module and that the KS domain should sit as the C-terminal domain.[82] In addition, KS substrate selectivity was shown by Jenner and co-workers who used intact protein mass spectrometry and a range of short polyketide-like mimics to investigate the transfer from upstream ACP to downstream KS. They concluded that KS domains clade according to the functionalisation of the substrate received from the upstream ACP.[84] This results in an updated module numbering and the biosynthetic pathway shown in Figure 10 uses this new nomenclature. Finally, the original analysis and publication by Lavigne *et al.* did not identify an ECH domain located downstream of the di-domain ACP3/4 within Bat3.[78] This additional domain is discussed in detail in Chapter 2.1.1 and included in the updated pathway.

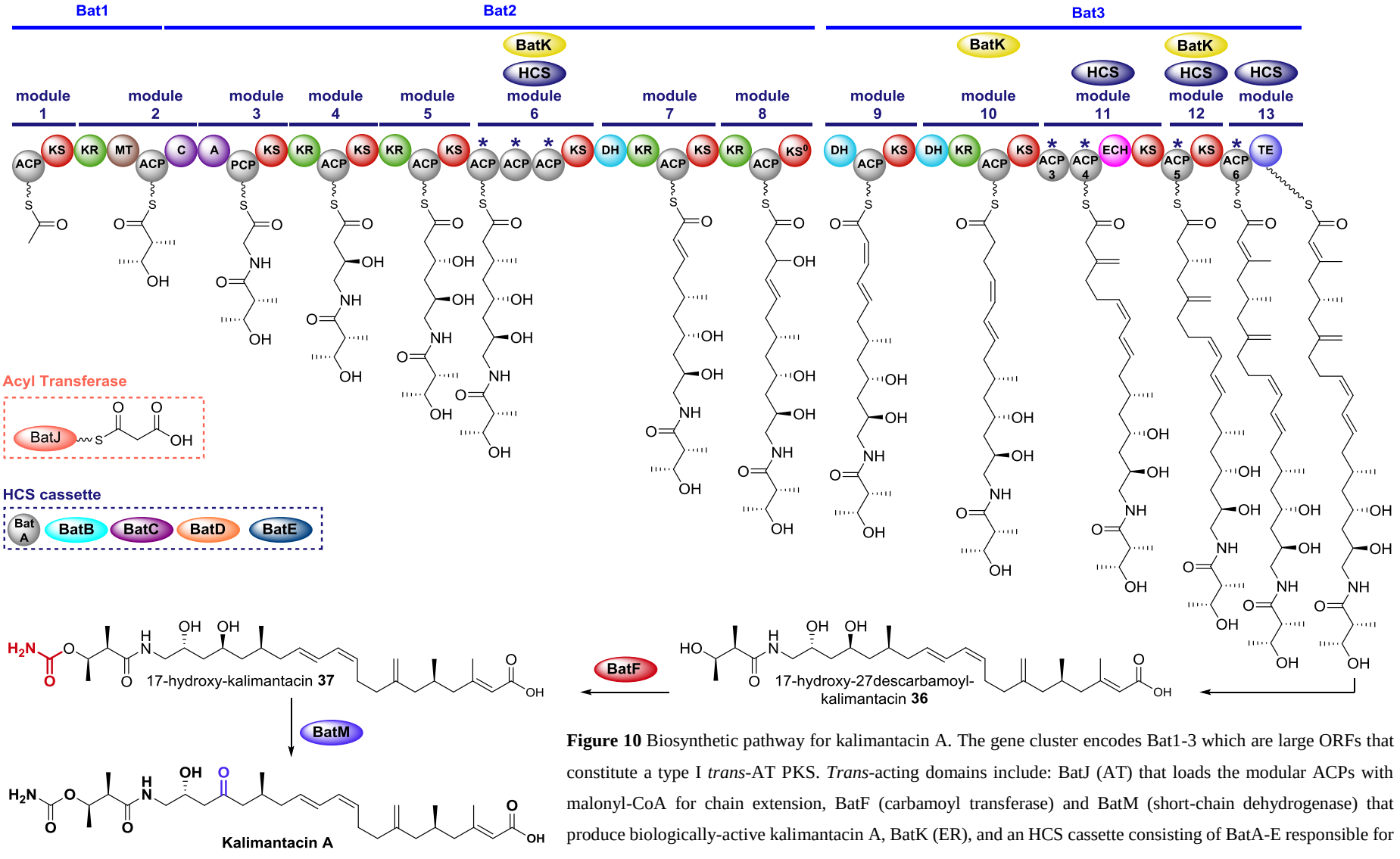


Figure 10 Biosynthetic pathway for kalimantacin A. The gene cluster encodes Bat1-3 which are large ORFs that constitute a type I *trans*-AT PKS. *Trans*-acting domains include: BatJ (AT) that loads the modular ACPs with malonyl-CoA for chain extension, BatF (carbamoyl transferase) and BatM (short-chain dehydrogenase) that produce biologically-active kalimantacin A, BatK (ER), and an HCS cassette consisting of BatA-E responsible for β-branching.

In the biosynthesis of kalimantacin, 11 malonyl derived building blocks are incorporated into the backbone by the type I PKS, along with a glycine incorporated by an NRPS module. A series of *trans*-acting enzymes further modify the backbone. Firstly, the fully reduced carbon backbone introduced in module 10 is achieved by the action of a *trans*-acting ER domain (Bat K). Following cleavage from the PKS to give **36**, BatF installs the carbamoyl functionality to the alcohol at C-27 to give **37** and BatM oxidises the C-17 alcohol to give kalimantacin A. Additionally, a series of four β-methyl branches, comprising three different types are incorporated in the polyketide structure by a *trans*-acting HCS cassette. The first branch is incorporated at Bat2 module 6 where a sat.-β-methyl group is introduced at C-15. The remaining three branches are installed consecutively in Bat3 modules 11, 12 and 13; an *exo*-β-methyl, sat-β-methyl and an *endo*-β-methyl respectively. Lavigne does not elude as to how the double bond isomers arise or how the reduction to produce the saturated methyl group occurs.[78]

Working in our lab, Dr Shushan Gao fed the producing *P. fluorescens* strain with [1,2-$^{13}C_2$]-acetate and determined the isotopic incorporation into kalimantacin A (Figure 11). Consistent with the proposed biosynthetic pathway, ^{13}C NMR analysis revealed 11 intact acetate units in the linear backbone, along with four β-methyl branches at C-3, C-5, C-7 and C-15 derived from [2-^{13}C] of the acetate. In a separate feeding experiment, L-[methyl-^{13}C]-methionine was shown to be the source of the C-29 α-methyl branch. The absence of ^{13}C-enrichment at C-19 and C-20 was consistent with the incorporation of glycine by the NRPS module.

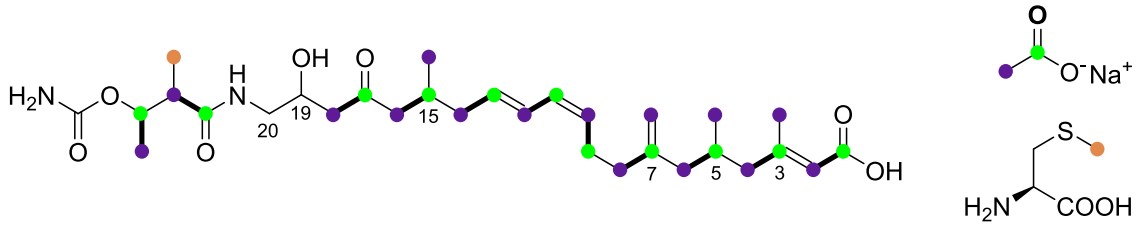


Figure 11 Isotopic labelling pattern for kalimantacin A obtained by feeding [1,2-$^{13}C_2$]-acetate and L-[methyl-^{13}C]-methionine to the wild-type *P. fluorescens* strain.

17-Hydroxy kalimantacin **37** was isolated from the knock-out strain *ΔBatM* arising from the lack of final oxidation (Figure 12). The *ΔBatF* strain is unable to catalyse carbamoylation and resulted in the isolation of two kalimantacin analogues; 17-hydroxy-27-descarbamoyl kalimantacin **36** and 27-descarbamoyl kalimantacin **38**. The loss of the carbamoyl group leads to a 10-fold reduction in antibiotic activity, whilst the 17-

hydroxylated compounds showed a complete loss of activity, demonstrating these final post-PKS tailoring steps are vital to producing the active compound.[78]

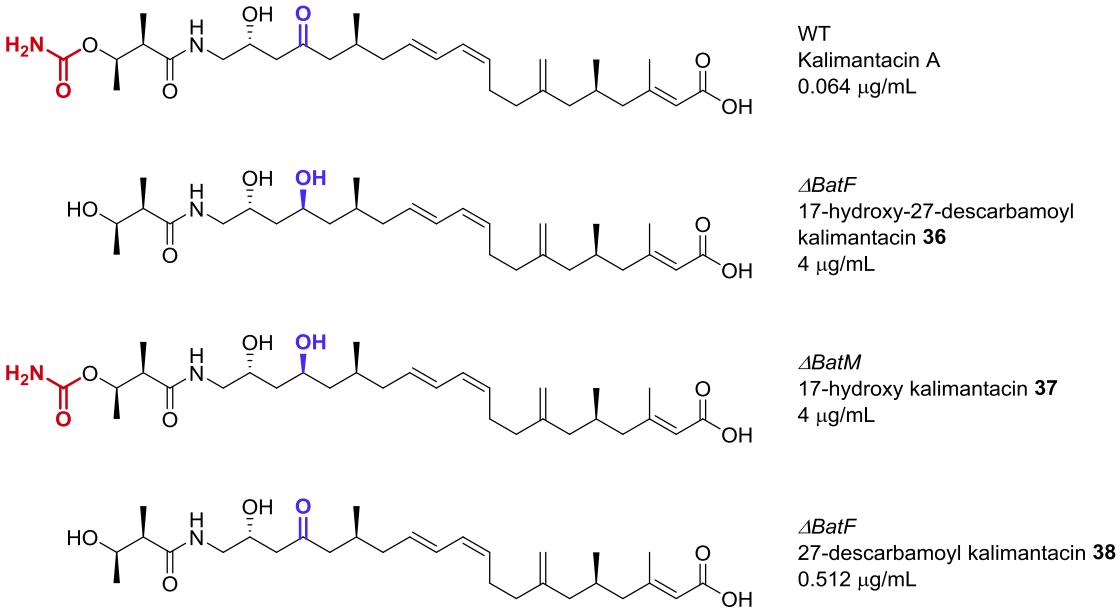


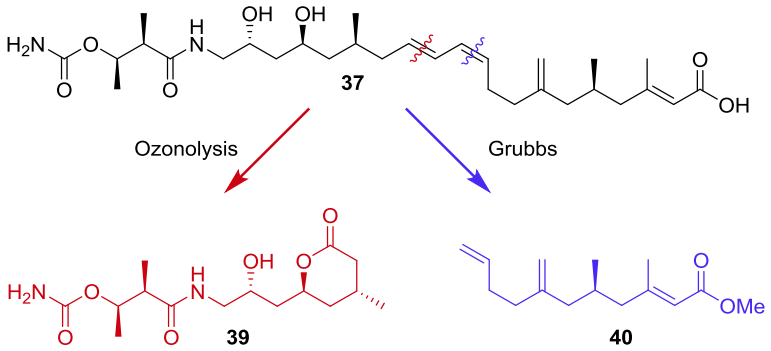
Figure 12 Bioactivity (MIC) of kalimantacin A and its analogues (**36**-**38**) isolated from WT and two knock-out strains (*ΔBatM* and *ΔBatF*) against S. *aureus* ATCC6538.

The kalimantacin gene cluster encodes two self-protection mechanisms: secretion of an inactive precursor prior to activation upon export and the presence of a FabI isoform. Kalimantacin is an inhibitor of FabI, the ER domain from type II fatty acid biosynthesis in bacteria.[85,86] The proposed mechanism of action has been debated in the literature, initiated by Klochko and co-workers who proposed that kalimantacin, analogous to mupirocin, acts as an inhibitor of aminoacyl tRNA synthetase.[80] This work was based on *in silico* modelling and an apparent similarity between the two molecular structures. This hypothesis was strongly rejected by O'Neill *et al.* who elegantly demonstrated the *in vivo* antibiotic target of activity of kalimantacin and mupirocin to be fatty acid and protein biosynthesis respectively.[87] This work supports the original hypothesis proposed by Lavigne *et al*.[88] The kalimantacin-producing *P. fluorescens* strain contains a FabI isoform (BatG) which can complement for FabI and confers full resistance to the producing bacteria but is non-essential for kalimantacin production. There has been much debate in the literature about the validity of FASII inhibitors in some strains, however, it remains a valid target for the clinically relevant S*taphylococcus aureus*.[89–91]

Export-coupled activation of an inactive intermediate prior to conversion to the toxic antibiotic is common in Nature.[92] Kalimantacin biosynthesis encodes late carbamoylation

and hydroxyl oxidation steps to produce the active antibiotic. Carbamoylation as the final, activation step has been shown for several secondary metabolites, including novobiocin that was 100-fold less active without the carbamoylation.[93,94] An additional advantage of the final, activating C-17 oxidation is its reversal of the reduction that takes place in Bat 2 module 5. A subtle difference between more common inactivation mechanisms (*O*-phosphorylation and *N*-acetylation) is displayed by this reaction. Resistance often occurs through horizontal gene transfer of antibiotic-inactivating enzymes.[95] However, the inactive 17-hydroxyl is introduced by a modular PKS, a step that is difficult to acquire through gene transfer, and reduces the chance of resistance.

In 2017, researchers in our lab reported the elucidation of the five previously unassigned stereocenters of kalimantacin A and the sixth stereocenter of the diol precursor (Scheme 13).[96–98] This was achieved through isolation of 17-hydroxy kalimantacin **37** and degradation by ozonolysis and Grubbs metathesis to fragments **39** and **40**. The total synthesis of both fragments was completed and their data compared with the corresponding natural product fragments to methodically determine each stereocenter.



Scheme 13 Elucidated stereochemistry of 17-hydroxy kalimantacin **37** and fragmentations.

1.6 Project Aims

Understanding the mechanisms and selectivity of enzymatic steps within biosynthetic pathways may allow for the rational engineering to produce novel compounds. The elucidation of the relative and absolute stereochemistry of kalimantacin laid the foundations for further work on its biosynthetic pathway to be undertaken. Kalimantacin contains four β-methyl branches that consist of three different types: an *endo*-β-methyl, an *exo*-β-methyl and two saturated β-methyl branches. This unprecedented arrangement of polyketide β-branches presented an opportunity to study the selectivity required by the biosynthetic machinery to incorporate multiple, different β-methyl branches. Our aim was to understand the selectivity of the HCS cassette in the consecutive installation of an *exo*-β-methyl at C-7, a saturated β-methyl at C-5 and an *endo*-β-methyl at C-3. This would require the synergy of synthetic chemistry for the preparation of biosynthetic mimics, biochemical methods for the expression, purification and characterisation of the HCS cassette and tailoring proteins, and the development of analytical techniques to monitor the enzymatic reactions.

2 Kalimantacin β-Branching Pathway

2.1 *In Vitro* Reconstitution of the HCS Cassette

At the outset of this project it was unclear how a single *trans*-acting HCS cassette in the kalimantacin biosynthetic pathway selectively installed the four β-branches. Whilst each type of β-branch has been reported in the literature, there are few compounds that include multiple, different β-branches that would require biosynthetic control, and none that contain three different types of β-branch.

2.1.1 Identification of a missing mECH domain

Investigations began with a BLAST search of BatE, the *trans*-acting ECH_2 from the HCS cassette, and a hit was identified for Bat3 with 53.9% amino acid sequence identity to BatE (Figure 13A). Bat3 is the third and largest ORF of the kalimantacin type I PKS that contains four elongating modules (9-12) and an off-loading module (13). The original assignment of Bat3 by Lavigne *et al.* did not annotate a domain that would be expected to show similarity to BatE.[78] However, re-analysis of the 328 amino acid sequence linking ACP4 and the downstream KS identified a potential modular ECH (mECH) domain.

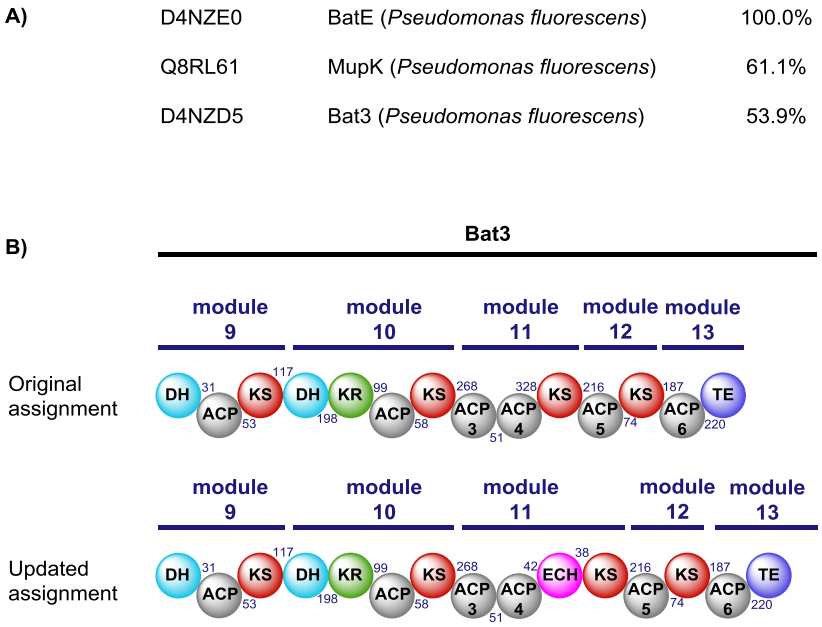
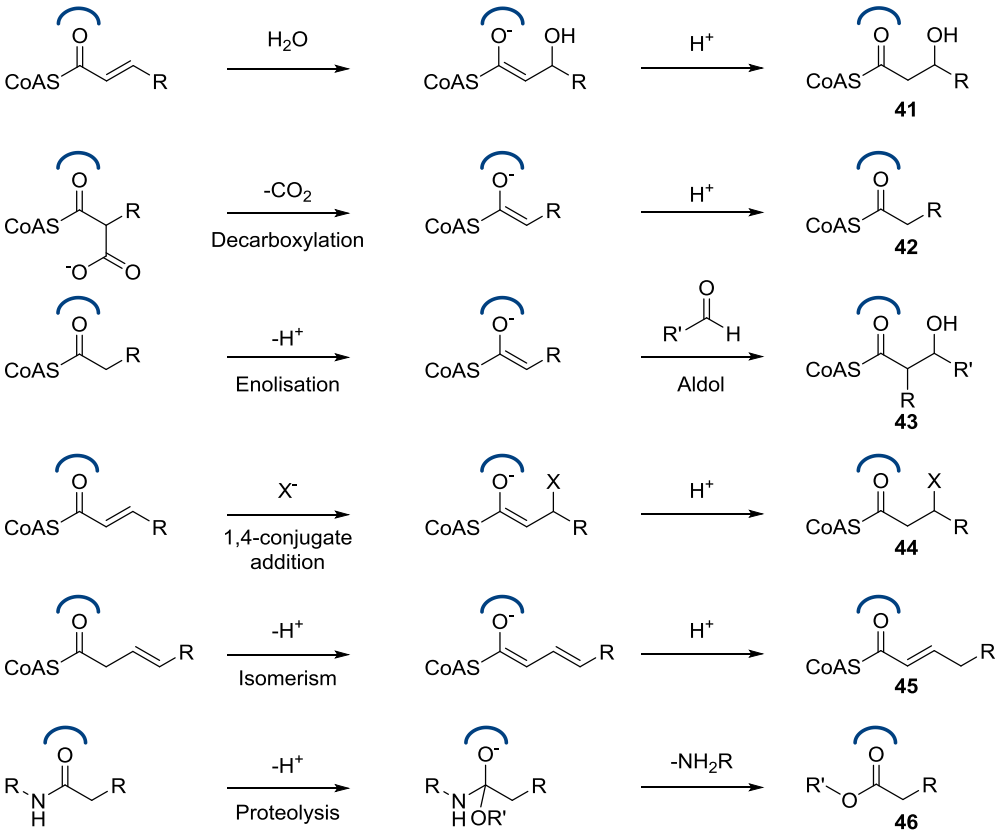


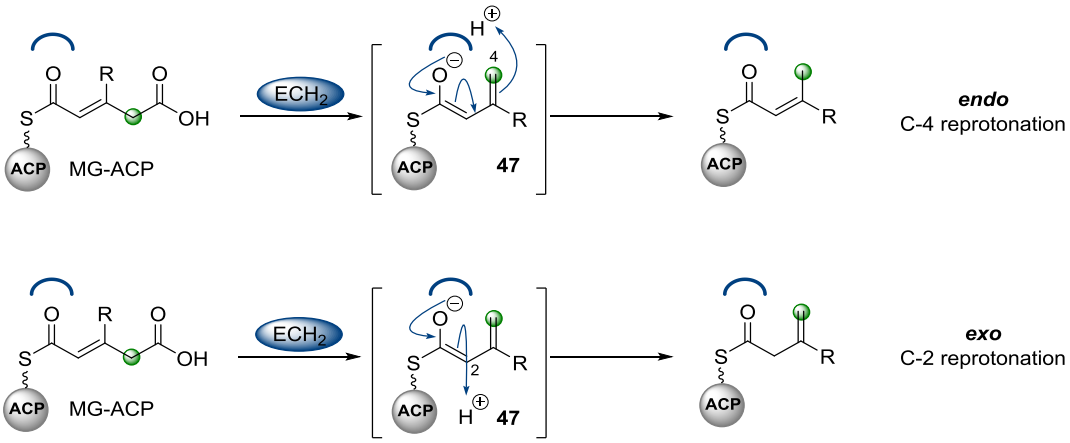
Figure 13 A) Extract from a BLAST search using BatE as a search query. An unexpected result was the sequence similarity to Bat3 which has previously not had an assigned ECH domain. B) Domain order of the third large ORF Bat3 as assigned by Lavigne *et al.*[78] The updated domain order to include the previously unassigned ECH domain with the linker length between catalytic domains shown in blue.

Secondary structure prediction (JPred4.0) and homology modelling using the I-Tasser/Phyre2 servers also suggested structural similarity between BatE and mECH. The two domains, along with BatD, belong to the enoyl-CoA hydratase (ECH) subgroup of the crotonase superfamily (CS) of enzymes.[99–102] Despite low sequence homology, this superfamily share a common structural scaffold, binding modes and mechanistic elements to catalyse a wide range of reactions. These include (de)hydration (**41,** BatD), (de)carboxylation (**42**, BatE), C-C/C-X bond formation (**43**-**44**), alkene/diene isomerism (**45**) and hydrolysis (**46**) (Scheme 14).[99,100] The shared feature in each mechanism is the formation of an oxyanion hole (OAH), in which two backbone NHs hydrogen bond to the carbonyl oxygen. Most commonly, crotonase reactions utilise thioester-bound CoA or ACP substrates resulting in the activation of a carbonyl group by polarisation of the C=O double bond. The electrophilic carbon atom is now more susceptible to nucleophilic attack or enolisation. The resulting enolate/oxyanion intermediate is also stabilised by the OAH, analogous to the stabilisation seen in the serine and cysteine proteases.[103,104]



Scheme 14 The range of reactions catalysed by the crotonase superfamily. CoA-thioester bound substrates are most common and the shared feature of catalysis is stabilisation of an enolate intermediate in an oxyanion hole (blue). Figure adapted from Schofield *et al*.[99]

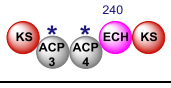
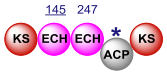
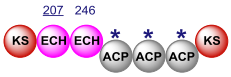
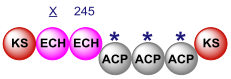
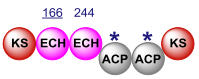
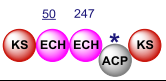
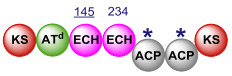
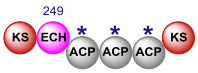
Considering the close homology of mECH with BatE and other well characterised ECH_2 domains from polyketide β-branching pathways, it is possible that mECH is also a decarboxylase. Its position in the modular PKS after ACP4 and before the elongating KS of the downstream module, suggests that it plays a role in the formation of the *exo*-β-methyl branch. As introduced earlier, the *cis*-AT PKS of jamaicamide incorporates an *exo*-β-methyl arising from decarboxylation of MG-ACP and C-2 reprotonation of intermediate **47** (Scheme 15).



Scheme 15 ECH_2-catalysed formation of *endo*- and *exo*-β-methyl branches. Enolate **47** is stabilised in an OAH (blue) and selective reprotonation at the C-2 or C-4 carbon results in formation of the *endo*- or *exo*-β-branch respectively.

Analysis of the modular PKS of *trans*-AT polyketides that possess HCS cassette derived *exo*-β-methyl branches revealed a common modular architecture whereby one or two ECH domains lie before the β-branching ACPs (Figure 14). Interestingly, the ACPs and ECH domain in the kalimantacin ORF are reversed in order, however, due to the many nuances associated with the 3D architecture of the modular PKS it is not thought to be significant in determining the β-branching outcome. The common architecture makes it likely that these ECH domains play a key role in *exo*-β-methyl formation and the assembly lines were probed further.

Inspection of the HCS cassette of pederin and the related compounds showed that they all contained *in-cis* ECH_2 domains, with the same general architecture (KS-ECH-ECH-ACP).[20,27] Most of the pathways are annotated with two modular ECH domains, however, closer inspection shows that the N-terminal domains are truncated, show low sequence homology to other ECH_2 domains and are likely to be non-functional. This would mean there is only a single, full length ECH_2 domain in each of the gene clusters. The number of ACPs varies between 1 to 3 copies, which is not uncommon for β- branching ACPs.

			ACP	KS^0	HMGS	ECH_1	ECH_2	Total number of ECH_2 domains	Total number of β-branches
A)	Kalimantacin Bat3 C-7 *exo*		BatA	BatB	BatC	BatD	BatE mECH	2	4
B)	Pederin PedI		PedN	PedM	PedP	PedL	PedI	1	1
	Onnamide OnnB		?	?	OnnA	?	OnnB	1	1
	Diaphorin DipP		DipF	DipR	DipE	DipC	DipP	1	1
	Nosperin NspA		NspG	NspH	NspI	NspJ	NspA	1	1
	Psymberin PsyA		?	?	PsyI	PsyJ	PsyA	1	1
C)	Phormidolide PhmE C-31 *exo*		PhmB	PhmA	PhmP	PhmN	PhmO PhmE PhmI	3	5
	Phormidolide PhmI C-9 *exo*								
	Spliceostatin fr9GH		Fr9M	Fr9N	FR9K	Fr9L	Fr9GH	1	1

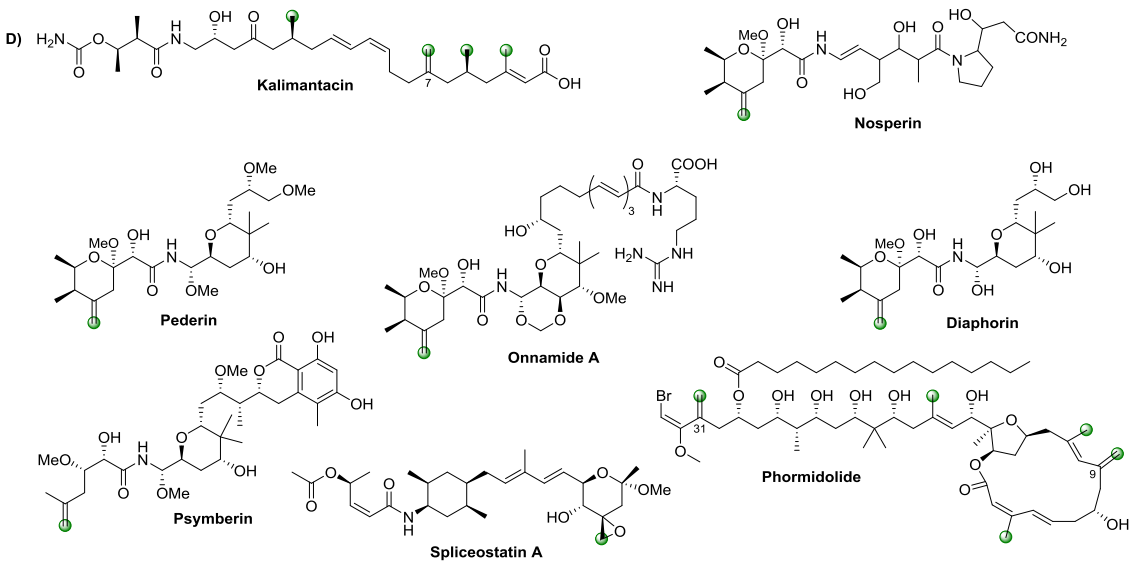
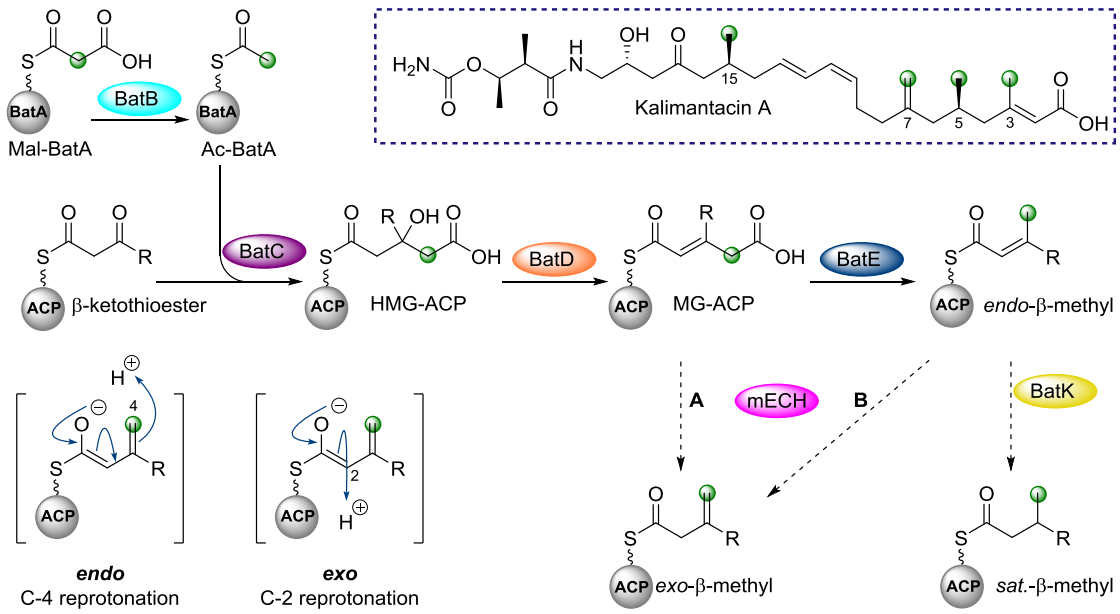


Figure 14 Modular architecture for the introduction of *exo*-β-methyl branches: kalimantacin (A), pederin-related compounds (B) and additional *exo*-β-methyl branch compounds (C). The ECH di-domain in the pederin-related compounds and phormidolide possess an N-terminal domain (underlined) that is often truncated and possibly non-functional. Domain length is annotated for ECH domains except DipP for which the length was not determined in the literature and sequence alignment was inconclusive for boundary assignment. ECH_2 domains shown in navy blue are *trans*-acting and those in pink are modular. D) Structures of the *exo*-β-methyl containing compounds. The *exo*-β-methyl in spliceostatin is subsequently epoxidised to give the final product.

Spliceostatin is another example that contains modular ECH domains and a *trans*-acting HCS cassette does not possess a free-standing ECH_2. Once again, the modular ECH must act in-*cis* to produce the *exo*-β-methyl branch.[105] In these examples, the presence of a single ECH_2 domain suggests the *exo*-β-methyl branch must be derived from a modular ECH_2-catalysed decarboxylation followed by C-2 reprotonation, analogous to the mechanistic steps reported for jamaicamide.

The gene cluster that produces phormidolide encodes the incorporation of five β-methyl branches; two *exo*-β-methyls and three *endo*-β-methyls.[68] A full *trans*-acting HCS cassette is present in addition to modular ECH domains (PhmE and PhmI) at the point of *exo*-β-methyl incorporation. Both of these modular ECH domains show sequence similarity to characterised ECH_2 (decarboxylase) domains but their biochemical role was not confirmed by the authors.[68] However, it is striking that the modular ECH domains are only present at the point of *exo*-β-methyl incorporation suggesting a common role in the formation of this moiety. This may be through decarboxylation/C-2 reprotonation, or alternatively isomerism of an *endo*-β-methyl branch formed by the *trans*-acting HCS cassette. Other than kalimantacin, this is the only example of a gene cluster that contains multiple ECH_2 domains and, due to the diverse range of catalysis of the CS, it is possible that the modular ECH may catalyse either decarboxylation or isomerism.

Consideration of these previously characterised pathways led to the following hypothesis for the incorporation of the four β-methyl branches in kalimantacin. All four branches share the same BatC-catalysed condensation step between Ac-BatA and a modular ACP_A-bound β-ketothioester to form HMG-ACP_A (Scheme 16). BatD-catalysed dehydration gives MG-ACP_A, followed by decarboxylation and C-4 reprotonation by BatE furnishes three *endo*-β-methyl branches (C-3, C-5 and C-15, kalimantacin numbering) as the modular architecture is identical for the domains associated with this incorporation. The two sat.-β-methyl branches (C-5 and C-15) arise from the subsequent selective reduction of the *endo*-β-methyl branches by the *trans*-acting ER BatK, which is likely to specifically recognise the cognate ACPs. Finally, the *exo*-β-methyl (C-7, kalimantacin numbering) may arise from one of two pathways. Firstly, BatD and BatE produce an *endo*-β-methyl branch which is then isomerised from the conjugated alkene to the *exo*-β-methyl by mECH (Scheme 16, pathway B). Alternatively, following BatD-catalysed dehydration mECH intercepts MG-ACP_A and catalyses decarboxylation and reprotonation at the C-2 position to give the *exo*-β-methyl (Scheme 16, pathway A).



Scheme 16 Proposed mechanism for the formation of β-methyl branches in the kalimantacin biosynthetic pathway. The *exo*-β-methyl may be made *via* one of two routes. Route A: Decarboxylation and C-2 reprotonation of MG-ACP by mECH. Route B: mECH-catalysed isomerism of an *endo*-β-methyl.

Sequence analysis and homology modelling did not reveal any information about the likelihood of the mECH domain to catalyse either of these reactions. The crotonase superfamily of enzymes catalyse a wide array of reactions, including both decarboxylation and isomerism and are notoriously difficult to functionally classify based on sequence or 3D architecture.[106,107] Single point mutations within enoyl-CoA isomerase (ECI) domains led to a change of function from isomerism to hydration, showing the subtlety in structure and function.[108]

Early literature reports for the isolation of *exo*-β-methyl containing natural products speculated that the moiety may result from isomerism of an *endo*-β-methyl.[109] This has since been shown not to be the case in jamaicamide and C-4 reprotonation must apply to the pederin-like compounds with only one ECH_2 domain in the gene cluster.[45] However, uncertainty still exists for the more complex systems, kalimantacin and phormidolide, that exhibit a full *trans*-acting HCS cassette as well as modular ECH domains that result in the incorporation of *endo*- and *exo*-β-methyl branches.

2.1.2 Purification of HCS cassette and ACPs

His_6-tagged BatA was expressed using a standard procedure, whereby 2 L of LB medium were inoculated and grown to OD_{600} 1.0, induced with 250 µM IPTG and incubated for 16 h at 16 °C. The cells were removed by centrifugation and the soluble fraction purified

by Ni^{2+}- affinity chromatography and buffer exchanged into reaction buffer (50 mM Tris, 100 mM NaCl, pH 8.0). For long-term storage, BatA was desalted into water and lyophilised and stored at -80 °C as a white solid. SDS-PAGE and MS were used to confirm the protein mass, which was consistent with apo-BatA (Figure 15).

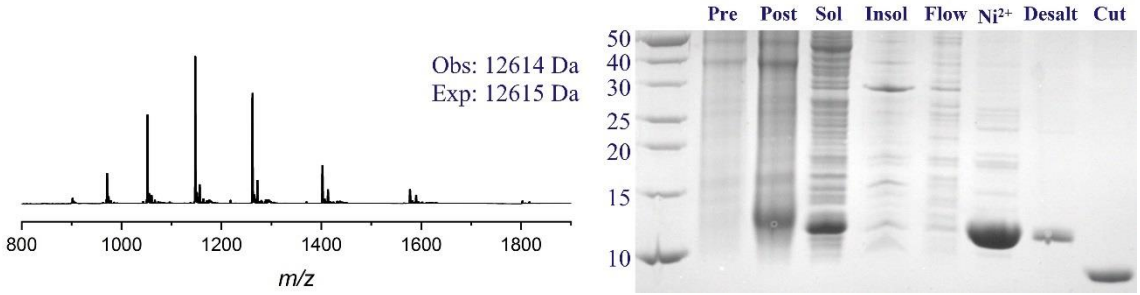


Figure 15 MS and SDS-PAGE analysis of apo-BatA. The second minor species is a Na^+ adduct of the desired protein. The His-tag was cleaved by TEV protease giving cut-BatA.

The pET151 expression vector contains a cleavage site (ENLYFQG) between the N-terminal His-tag and the desired protein sequences that allows for the removal of the His-tag by Tobacco Etch Virus (TEV) protease. The protease cleaves between glutamine (Q) and glycine (G) to remove the 27 amino acid his-tag containing region, leaving only a short 6 amino acid tag on the protein of interest (GIDPFT). Throughout this work, both cut- and uncut- BatA have been used with no noticeable difference in their behaviour.

To check the function of BatA, the inactive apo-BatA was converted into holo-BatA. The transfer reaction to form a phosphodiester bond to the ACP is catalysed by a phosphopantetheinyl transferase enzyme (PPTase) using CoA as the source of the phosphopantetheine (Ppant) arm and Mg^{2+} as a co-factor. This superfamily of enzymes has been categorised into group I (AcpS-type) and group II (Sfp-type) PPTases. Helix II of the ACP is the key interaction site for both groups of PPTases with the AcpS-type interacting through electrostatics, whilst hydrophobic interactions are key for Sfp-type PPTases.[110,111] AcpS-type PPTases show narrow-range selectivity for ACPs, mainly active with type II fatty acid and polyketide ACPs. AcpS from *Streptomyces coelicolor* has been crystallised and its promiscuity in transferring a wide range of CoA thioesters demonstrated.[112,113] Sfp from *Bacillus subtilis* is the archetypal type II PPTase and is well characterised.[114–116] This group of PPTases are highly promiscuous towards both the ACP and CoA derivatives and are regularly used during the *in vitro* reconstitution of biosynthetic pathways.[110,111,117]

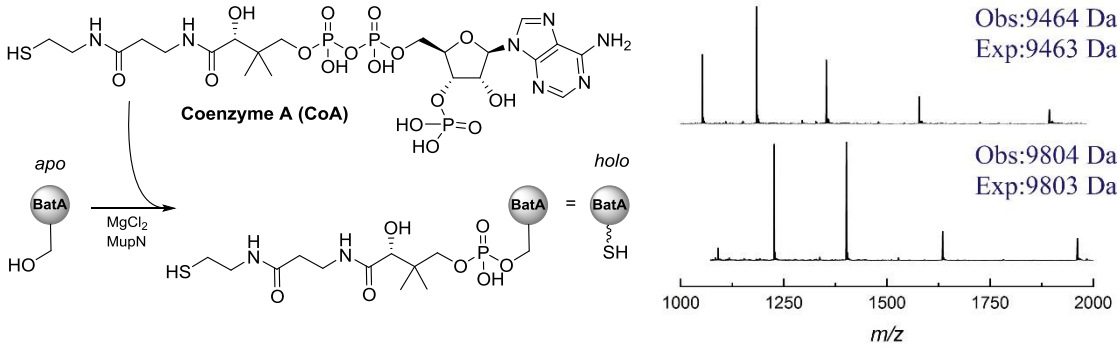


Figure 16 Activation of apo-BatA (9464 Da) by transfer of a Ppant arm from CoA to a conserved serine residue to give holo-BatA (9804 Da).

The conversion of apo- to holo- BatA was investigated using both types of PPTases: *E. coli* AcpS (ecAcpS) which is a type I PPTase, and MupN which is a type II PPTase from the mupirocin biosynthetic pathway. MupN has been shown to load both type I and type II mupirocin ACPs, so it was anticipated that it would act on ACPs from a non-cognate system.[118] Upon incubation with CoA, $MgCl_2$ and ecAcpS for 3 hours, apo-BatA did not undergo conversion to holo-BatA. However, incubation under the same conditions with MupN resulted in the conversion to holo-BatA after only one hour (Figure 16). The selectivity for a type II PPTase was expected and allowed for the activation of kalimantacin ACPs.

BatC is the HMGS that transfers acetate from Ac-BatA to a β-ketothioester attached to a modular ACP_A. Expression and purification using standard conditions yielded a protein of the correct mass by SDS-PAGE and MS (Figure 17). The concentrated protein (300 µM) was stored at -80 °C and was stable for two years.

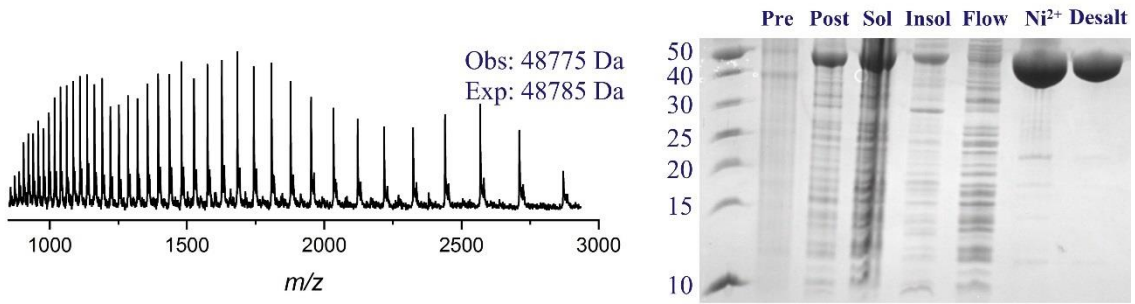


Figure 17 MS and SDS-PAGE characterisation of BatC.

BatD and BatE are both ECH domains that carry out the dehydration (BatD) and subsequent decarboxylation (BatE) of the HMG-ACP_A intermediate. Both were expressed as soluble proteins, purified by Ni^{2+}-chromatography and analysed by SDS-

PAGE and MS. Interestingly, BatD runs between 25 and 30 kDa by SDS-PAGE, however, the MS indicates a mass of 30402 Da, an error of only 2 Da (Figure 18A). BatE shows the same characteristic of running at a lower mass by SDS-PAGE, however, the mass recorded (31361 Da) was within 2 Da of the expected mass (Figure 18B).

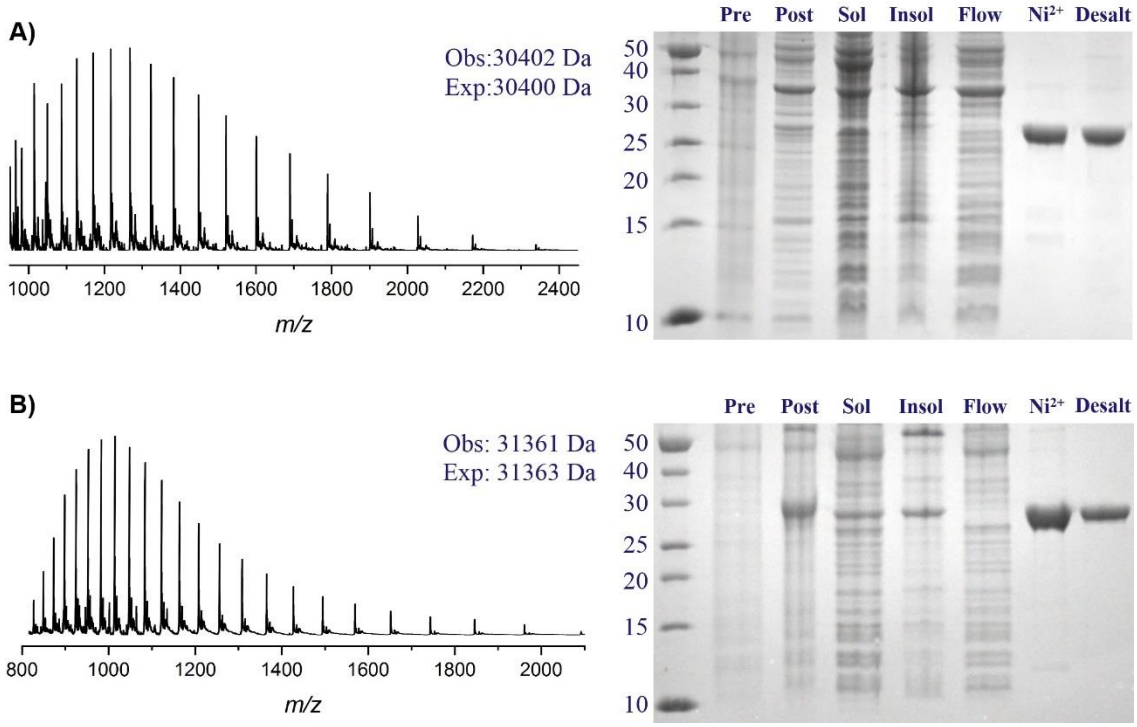


Figure 18 MS and SDS-PAGE analysis of A) BatD and B) BatE.

Analytical gel filtration using a pre-calibrated column was carried out on both BatD and BatE to determine their oligomeric state in solution. The retention time of BatD (13.36 mins) and BatE (13.42 mins) suggested both proteins were trimeric in solution. This was consistent with the reported association states of the crotonase superfamily that have been reported to be trimers or dimers of trimers (Figure 19).[100]

Bat3 is the large ORF which interacts with the HCS cassette to introduce the three consecutive β-methyl branches in kalimantacin. The di-domain ACP in module 11 (ACP3/4) bears the substrate for the *exo*-β-methyl incorporation, ACP5 (module 12) for the sat.-β-methyl and ACP6 (module 13) for the *endo*-β-methyl. Starting with the tandem ACP, the gene for ACP3/4 was synthesised and cloned into pET151. The plasmid was transformed, a pre-culture grown and used to inoculate LB medium. Following induction with 250 µM IPTG and 16 h incubation at 16 °C, the cells were harvested, and the soluble fraction purified by Ni^{2+}-affinity chromatography. A protein of the correct molecular mass was observed by SDS-PAGE, however, MS analysis revealed that the di-domain

ACP had been expressed in the holo form. This suggests that the ACP interacts with ACPS from *E. coli* and cellular CoA to enable PPT transfer.[119]

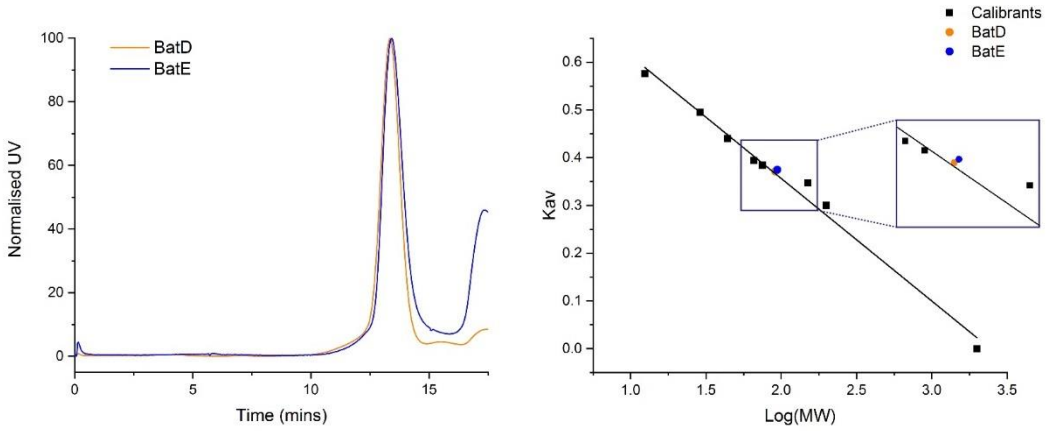


Figure 19 S200 analytical gel filtration trace of BatD (13.36 mins) and BatE (13.42 mins) to determine molecular weight/oligomeric state. The trimeric moleuclar weight of both BatD and BatE was used.

This modification was not ideal as the expression of holo-ACP prevents the subsequent loading of biosynthetic substrate mimics by transfer of a functionalised Ppant arm. It is possible to load ACPs by *trans*-thioesterification by incubation of holo-ACP with an excess of the desired substrate as an *N*-acetylcysteamine (SNAC) derivative.[81] However, this is an inefficient and non-specific reaction which may functionalise alternative nucleophilic sites on the ACP. There is currently no reliable method for the removal of a PPT arm. Work by Burkart and co-workers has identified AcpH from *E. coli* and homologs from other bacterial organisms that hydrolyse the Ppant arm.[120] However, the characterised proteins do not show the same level of promiscuity as the PPTases and this route was not attempted.

Production of only holo-ACP3/4 was not entirely consistent and initial purifications also yielded a mixture of apo/apo-ACP3/4 and holo/holo-ACP3/4. To remove the holo/holo-ACP3/4, the mixture was treated with 5 µM $CuCl_2$ and purified by size exclusion chromatography (Figure 20). The di-sulfide dimer could be separated from apo/apo-ACP3/4, however, the inefficient purification method resulted in a low yield of the desired protein and the method was no longer investigated.

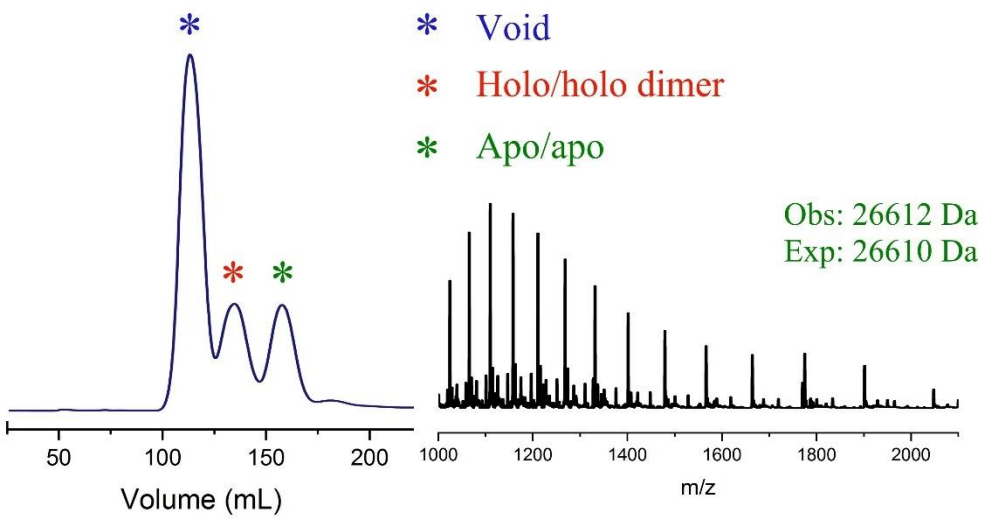


Figure 20 S75 size exclusion chromatography (SEC) trace separating the disulfide-bridged holo/holo-ACP3/4 dimer from apo/apo-ACP3/4 (26612 Da).

Inspection of the ACP sequences showed a 57% identity (82% similarity) between ACP3 and ACP4 and it is likely the function of the tandem domain is to increase flux through the PKS rather than any fundamental distinction (i.e. to deliver different substrates).[52,53] As a result, they were cloned as two single domain ACPs by point mutation (ACP3) or amplification/recombination (ACP4). Both ACPs were grown in LB, induced (250 µM IPTG) and incubated for 16 h at 16 °C, which furnished holo-ACP for both constructs. There was a significant difference in yield and stability of the obtained protein, with ACP3 yielding less than 0.5 mg/L with a tendency to precipitate from solution, whilst ACP4 was expressed around 6-7 mg/L. As a result, optimisation of expression was carried out only on ACP4, including using smaller flasks, different *E. coli* expression strains, IPTG concentration and expression time.

Using a standard protocol of inducing with 250 µM IPTG for 16 h gave only holo-ACP4. A time course with the same concentration of IPTG showed that after 2 or 4 h only apo-ACP4 was obtained and by 6 h, a small fraction of holo-ACP4 was observed. Decreasing the induction time reduced the protein yield so to compensate for this the concentration of IPTG was increased to 1 mM. Overall, the optimal conditions were deemed to be 1 mM IPTG induction for 5 h, yielding 7-8 mg/L apo-ACP4 (Figure 21).

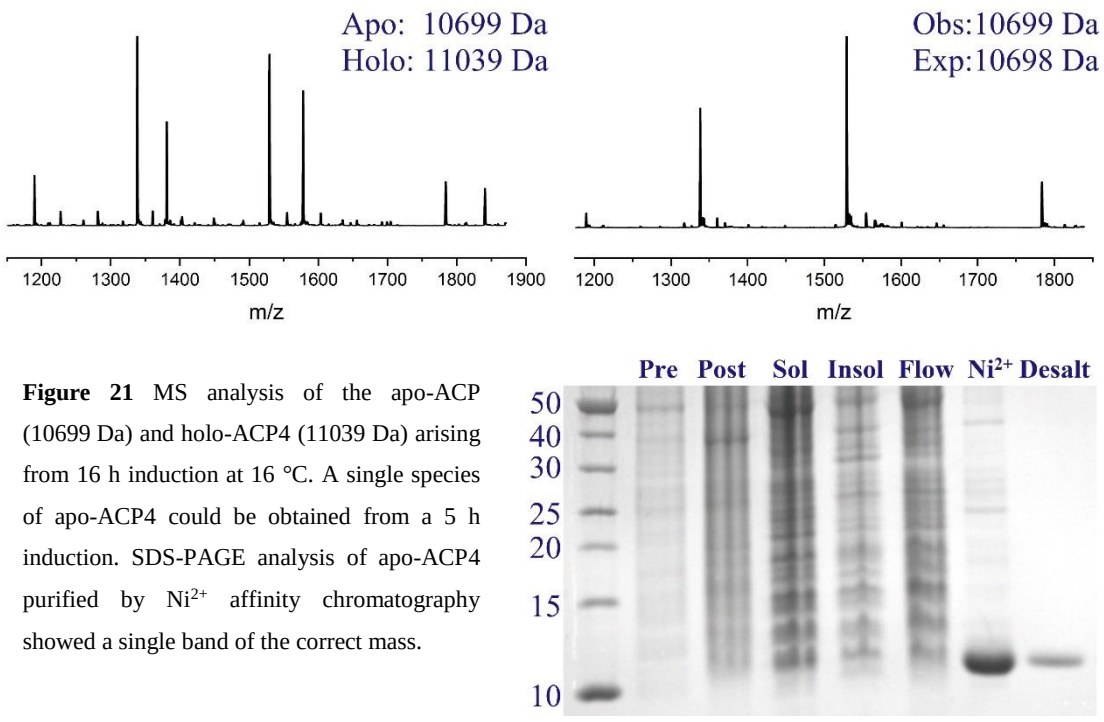


Figure 21 MS analysis of the apo-ACP (10699 Da) and holo-ACP4 (11039 Da) arising from 16 h induction at 16 °C. A single species of apo-ACP4 could be obtained from a 5 h induction. SDS-PAGE analysis of apo-ACP4 purified by Ni^{2+} affinity chromatography showed a single band of the correct mass.

ACP5 and ACP6 were grown in LB medium and induced with 250 µM IPTG for 16 h at 16 °C. Both proteins were identified in the soluble fraction, purified by Ni^{2+}-affinity chromatography and characterised by SDS-PAGE and MS (Figure 22). A single protein band consistent with the mass of apo-ACP was identified for both ACP5 and ACP6.

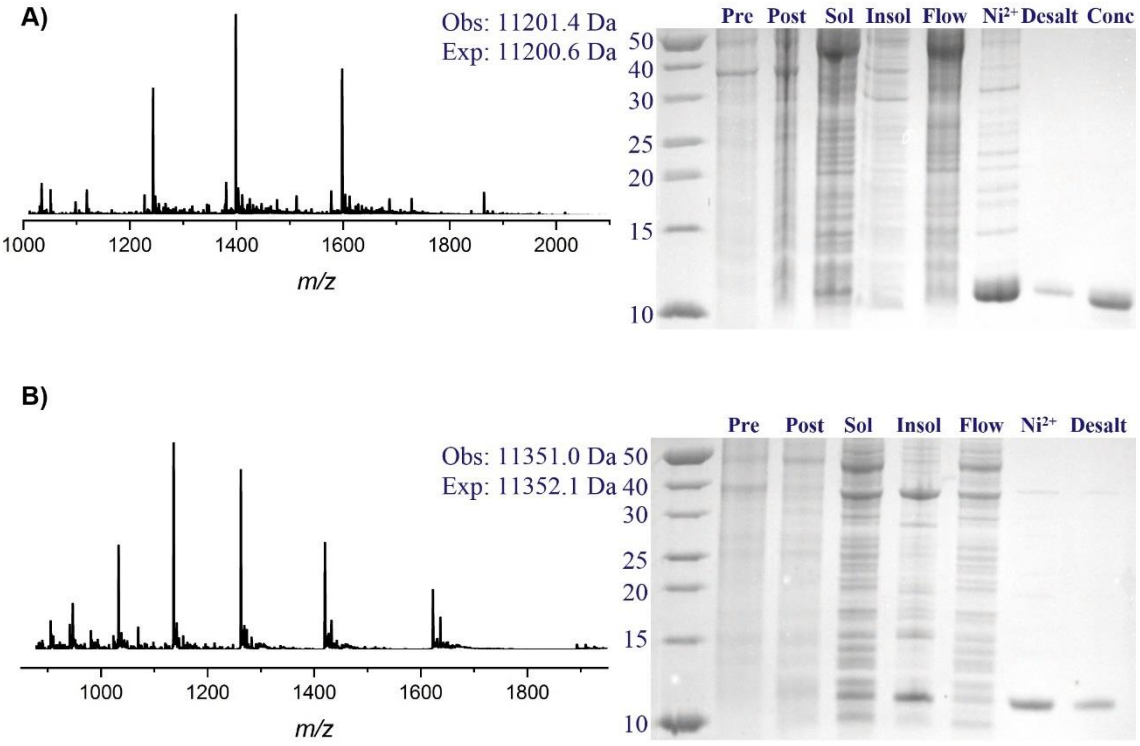


Figure 22 MS and SDS-PAGE characterisation of A) ACP5 and B) ACP6.

Subsequent cultures were grown with a 16 h induction time that did, however, occasionally result in formation of around 20% holo-ACP. ACP6 was poorly expressed when induced with 1 mM IPTG for 5 h so 10-12 h evolved as the optimal induction time. These results suggested a difference between the PPTase recognition motifs of ACP3/4 which are phosphopantetheinylated *in vivo* following an 16 h induction, compared with ACP5 and ACP6 which predominantly express as apo-ACP after the same time.

Following their expression, all ACPs were desalted into reaction buffer and kept at 4 °C for short-term storage. Aliquots of ACP4 and ACP5 in reaction buffer were stored at -20 °C and gently thawed on ice prior to assays. However, ACP6 was not stable to freeze-thaw conditions and precipitated after thawing on ice. As a result, ACP6 had to be freshly purified from a cell pellet prior to conducting assays.

2.1.3 Purification of tailoring domains

Module 10 catalyses a full round of ketoreduction, dehydration and enoyl reduction but lacks a dedicated *cis*-ER. BatK is hypothesised to act as a *trans*-ER which is recruited to this module to reduce the alkene moiety (Figure 23). It is also hypothesised that BatK acts on an *endo*-β-methyl branch attached to the ACP of module 6 and 12 to give the sat.-β-methyl branch. These three alkene reductions are the only examples in the kalimantacin biosynthetic pathway.

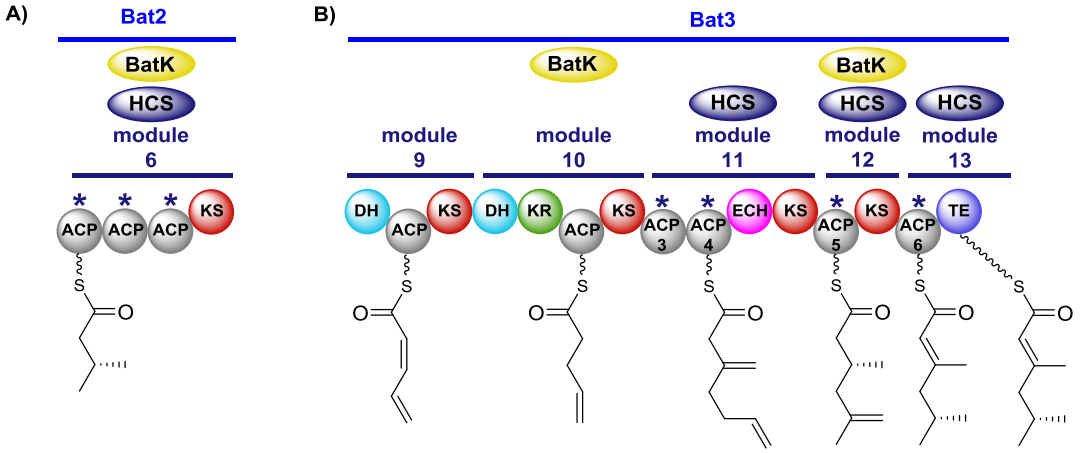


Figure 23 Extract of the kalimantacin biosynthetic pathway showing the proposed BatK-catalysed reductions. A) In module 6 of Bat2, a sat.-β-methyl is formed on the tri-domain of β-branching ACPs. B) Module 10 of Bat3 lacks an *in-cis* ER domain but full reductive processing gives an alkyl chain. ACP5 of module 12 tethers the intermediate for sat.-β-methyl formation.

The gene corresponding to BatK was cloned from genomic DNA and ligated into POPINF. The protein was expressed and purified by Ni^{2+}-affinity chromatography. The yellow, soluble protein was analysed by SDS-PAGE, which indicated a protein of the correct mass, however analysis by MS was not possible (Figure 24). By analytical gel filtration, the purified protein appeared to be monomeric, in contrast to the previously reported dimeric *trans*-ER DifA from the *trans*-AT PKS difficidin biosynthetic pathway.[121] The yellow colour of BatK suggested co-purification with a FAD/FMN co-factor.[122]

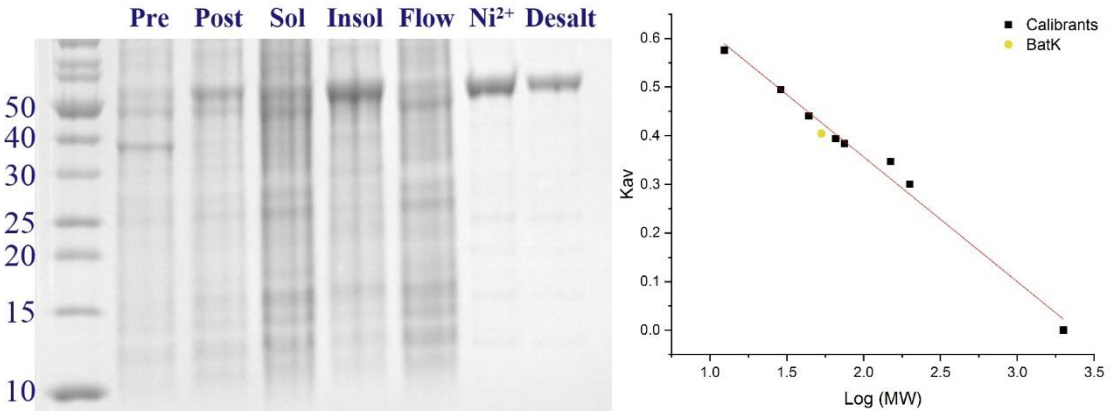


Figure 24 Characterisation of BatK by SDS-PAGE and S200 analytical size exclusion chromatography. The retention time was consistent with BatK being monomeric in solution.

Following the identification of the missing mECH domain, a construct was designed for recombinant expression. The domain boundaries were assigned by secondary structure prediction, homology modelling and comparison with literature crystal structures.[55] The gene was synthesised, cloned in pET151 and expressed. The soluble fraction was purified by Ni^{2+}-affinity chromatography and a protein of the correct expected mass was confirmed by SDS-PAGE and MS (Figure 25).

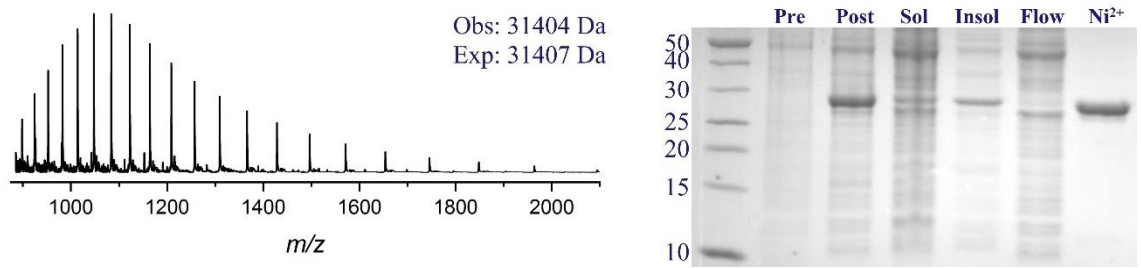


Figure 25 MS and SDS-PAGE analysis of mECH.

The expression of all the required protein components was the first step towards reconstituting the β-branching pathway *in vitro*. Demonstrating that each individual protein was functional required the preparation of surrogates for biosynthetic intermediates which would be loaded onto their respective carrier proteins.

2.1.4 Synthesis of pantetheines

A wide range of tools have been developed to study the enzymatic reactions of PKS pathways both *in vitro* and *in vivo* (Figure 26).[117] The gold standard is ACP-bound intermediates as the protein-protein interactions between the ACP and catalytic domain are often key to the correct delivery of substrate.[81] This method has been restricted by the high cost and limited choice of commercially available CoA derivatives.[123] Moreover, functionalisation of CoA by coupling to the thiol is challenging: yields are often poor and purification of the product is non-trivial due to their polarity.[124] More commonly, SNAC thioesters which mimic the end of a Ppant arm of an ACP are used due to their relatively straightforward synthetic accessibility.[84,125–127] This method generally requires incubation of the SNAC intermediates with the target protein, reaction quenching and isolation of the functionalised SNAC product. Consequently, a large quantity of SNAC derivative is required to monitor the reaction over sufficient time points coupled with the loss of material during work up.

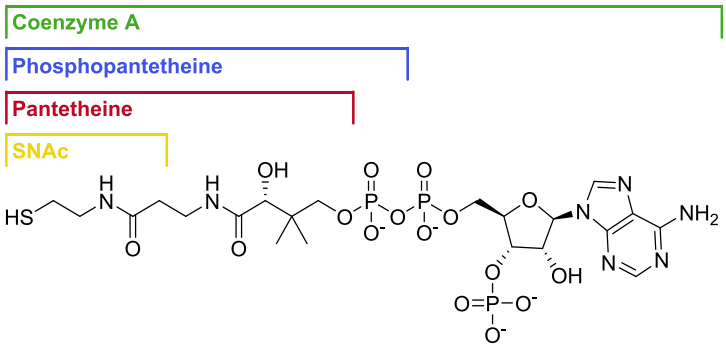
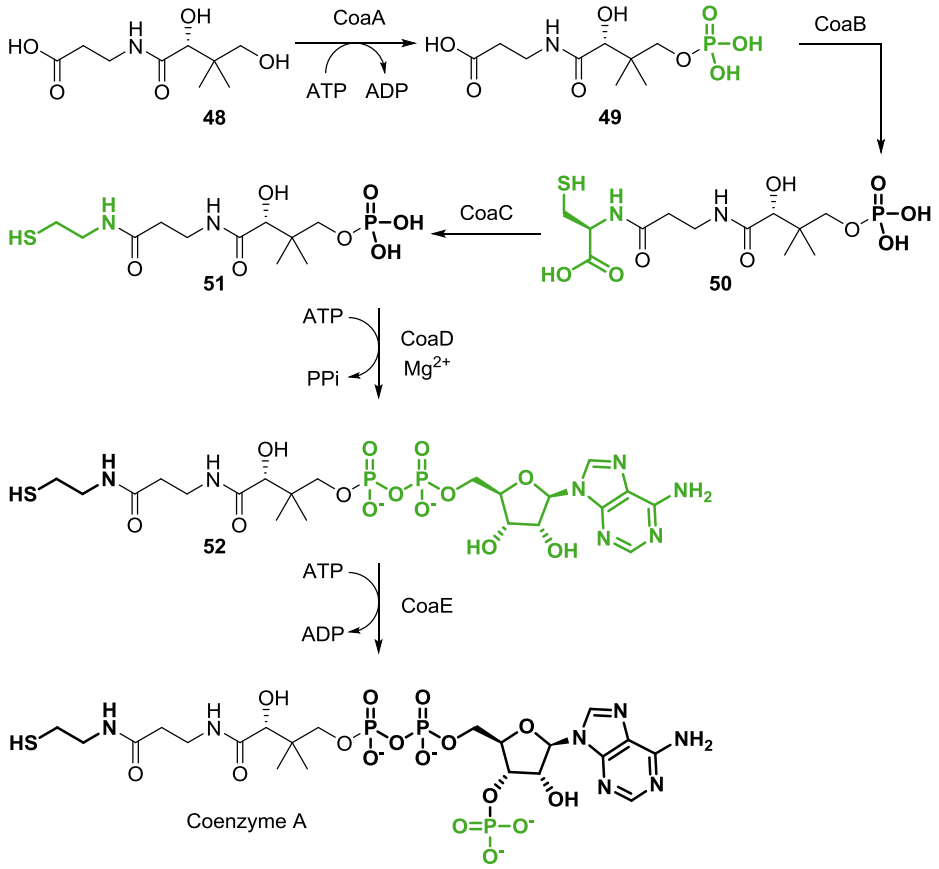


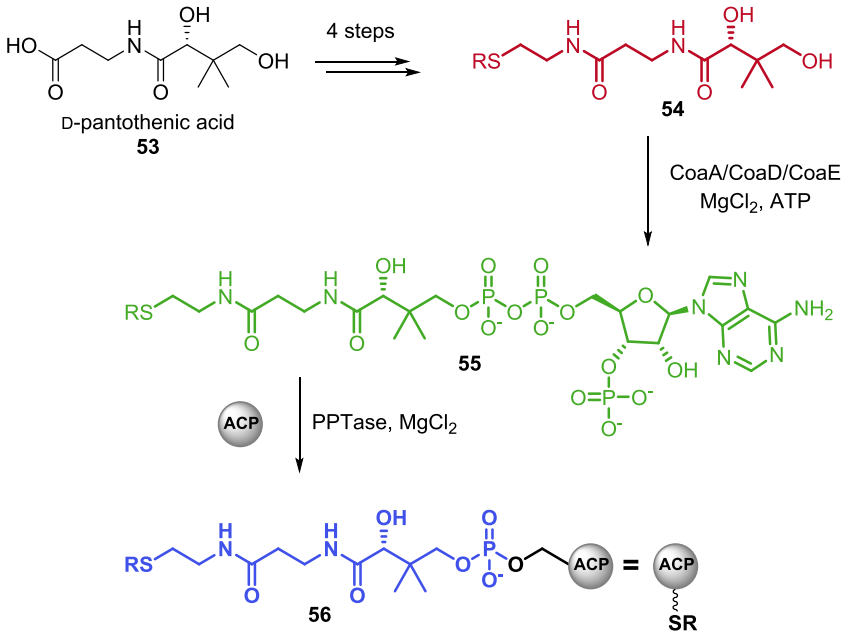
Figure 26 The structure of CoA and the structural mimics that are used to monitor enzymatic reactions.

Building on fundamental work on the CoA biosynthetic pathway, a new chemo-enzymatic method for the specific, cost-effective and synthetically amenable loading of ACPs has been developed.[128,129] The biosynthetic route to CoA is shown in Scheme 17, with the newly-added functional group of each catalytic step highlighted in green. CoaA and CoaE catalyse the synthetically challenging phosphorylation reactions and CoaD the adenylation step. CoaB and CoaC are responsible for the condensation and decarboxylation of cysteine to introduce the terminal thiol.



Scheme 17 CoA biosynthetic pathway with the new functional groups added by each enzyme shown in green.

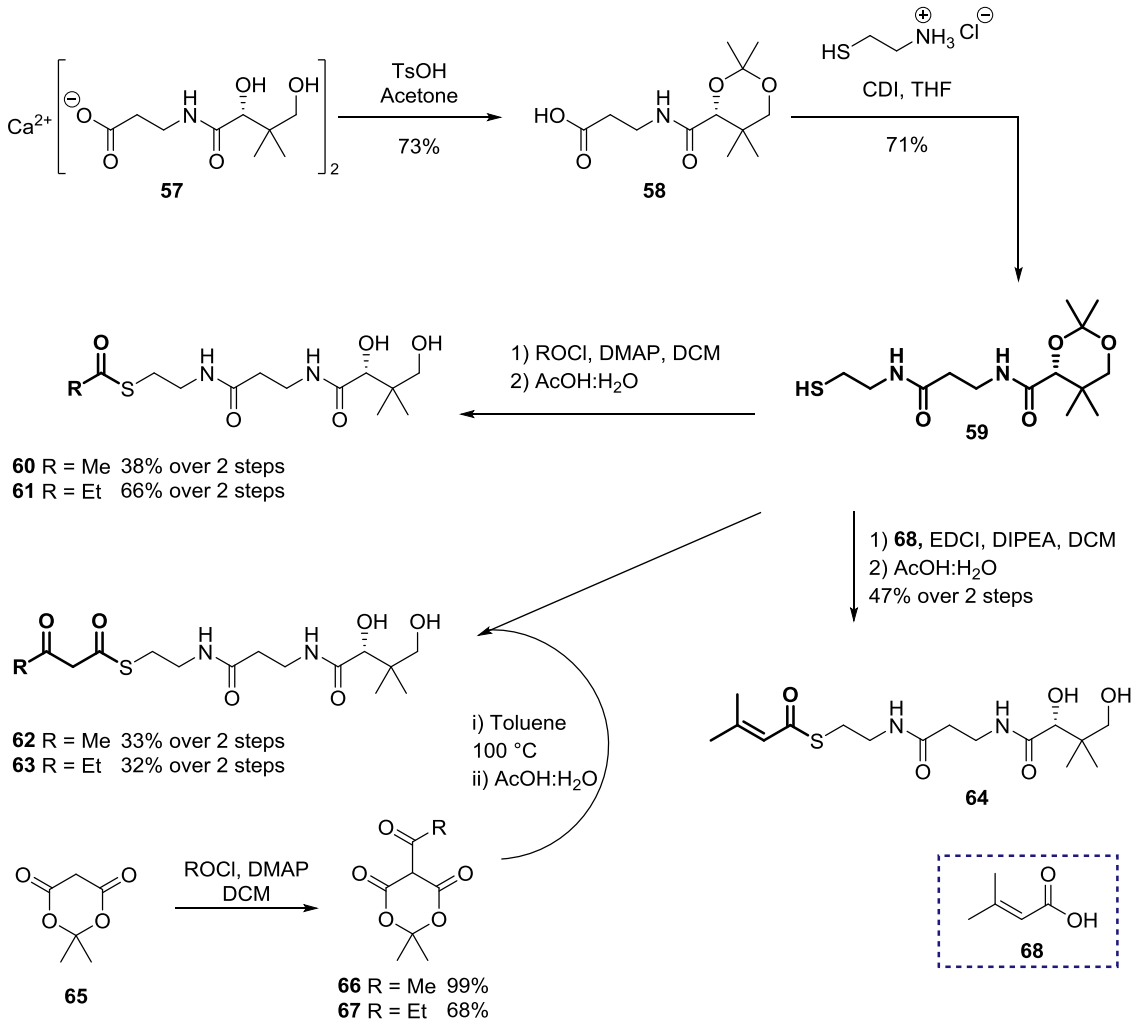
In the chemo-enzymatic method, synthetic chemistry is used to prepare a functionalised pantetheine **54**, with an amide coupling reaction replacing the CoaB and CoaC-catalysed steps. The terminal thiol may then be functionalised by thiol coupling reactions to give a range of acyl-pantetheine derivatives, which may be synthesised in four steps from a commercially available source of D-pantothenic acid **53** (Scheme 18).[124] These pantetheines may be used for enzymatic reactions due to the mimicry of the ACP-bound Ppant arm, but may also be sub-optimal.[130,131] Once synthesised, acyl-pantetheines may be "upgraded" into their CoA derivatives **55** by taking advantage of the promiscuity of the CoA biosynthetic machinery.[117,132] Using CoaA/CoaD/CoaE and the required co-factors (ATP and Mg^{2+}), the functionalised pantetheines are converted to the respective CoA **55**. A further extension of this methodology is the loading of ACPs with the functionalised CoA by adding the relevant PPTase into the upgrade reaction. This allows for the one-pot, chemoenzymatic loading of apo-ACPs to give functionalised ACPs **56** for the study of biosynthetic pathways.[132]



Scheme 18 Preparation of functionalised-ACP **56** in a one-pot, *in situ* chemoenzymatic pantetheine upgrade and ACP-loading reaction. D-pantothenic acid **53** is converted to pantetheine derivatives **54**, which are upgraded to derivatised CoA **55** and then the phosphopantetheine arm (blue) is transferred to apo-ACP by a PPTase to give functionalised ACP **56**.

In this project, the synthesis of derivatised pantetheine started with the protection of commercially available D-pantothenic acid hemicalcium salt **57** (Scheme 19). An amide coupling between acid **58** and cysteamine hydrochloride furnished the key protected pantetheine intermediate **59**. This compound was a stable, white solid that could be stored at ambient conditions prior to functionalising the thiol moiety. **59** was reacted with acetyl or propionyl chloride in DCM with DMAP to give the protected acyl-pantetheines. Deprotection with a 2:1 acetic acid/water mix gave acetyl-pantetheine **60** (38%) and propionyl-pantetheine **61** (66%) over the two steps.

The reaction of acetyl chloride or propionyl chloride with Meldrum's acid **65** led to the formation of acetyl-Meldrum's acid **66** and propionyl-Meldrum's acid **67**. These compounds were separately reacted with **59** in toluene at 100 °C to give acetoacetate (Acac)-pantetheine **62** and propionylacetate (PropAc)-pantetheine **63,** following deprotection of the acetal. Finally, synthesis of **64** was achieved *via* an EDCI-mediated coupling of **59** with acid **68** in THF. Deprotection under standard conditions gave **64** in 47% yield over the two steps.



Scheme 19 Synthesis of key protected pantetheine intermediate **59**. Further functionalisation furnished a range of acyl (**60-61, 64**) and β-ketothioester (**62-63**) pantetheines required for *in vitro* reconstitution of the β-branching pathway.

To test whether the CoA upgrade enzymes were functional, a preparative-scale conversion of Ac-pantetheine to Ac-CoA was trialled.[124] After 1 h, the reaction was quenched, and the reaction mixture analysed by LC/MS (Figure 27). Ac-pant ($[M-H]^-$ *m/z* 319 Da) was fully consumed, and a new peak corresponding to a more polar product was eluted from the column with *m/z* 808, corresponding to $[M-H]^-$ for acetyl-CoA. The product was purified on a preparative scale LC column and was confirmed as Ac-CoA by ^{1}H NMR spectroscopy.

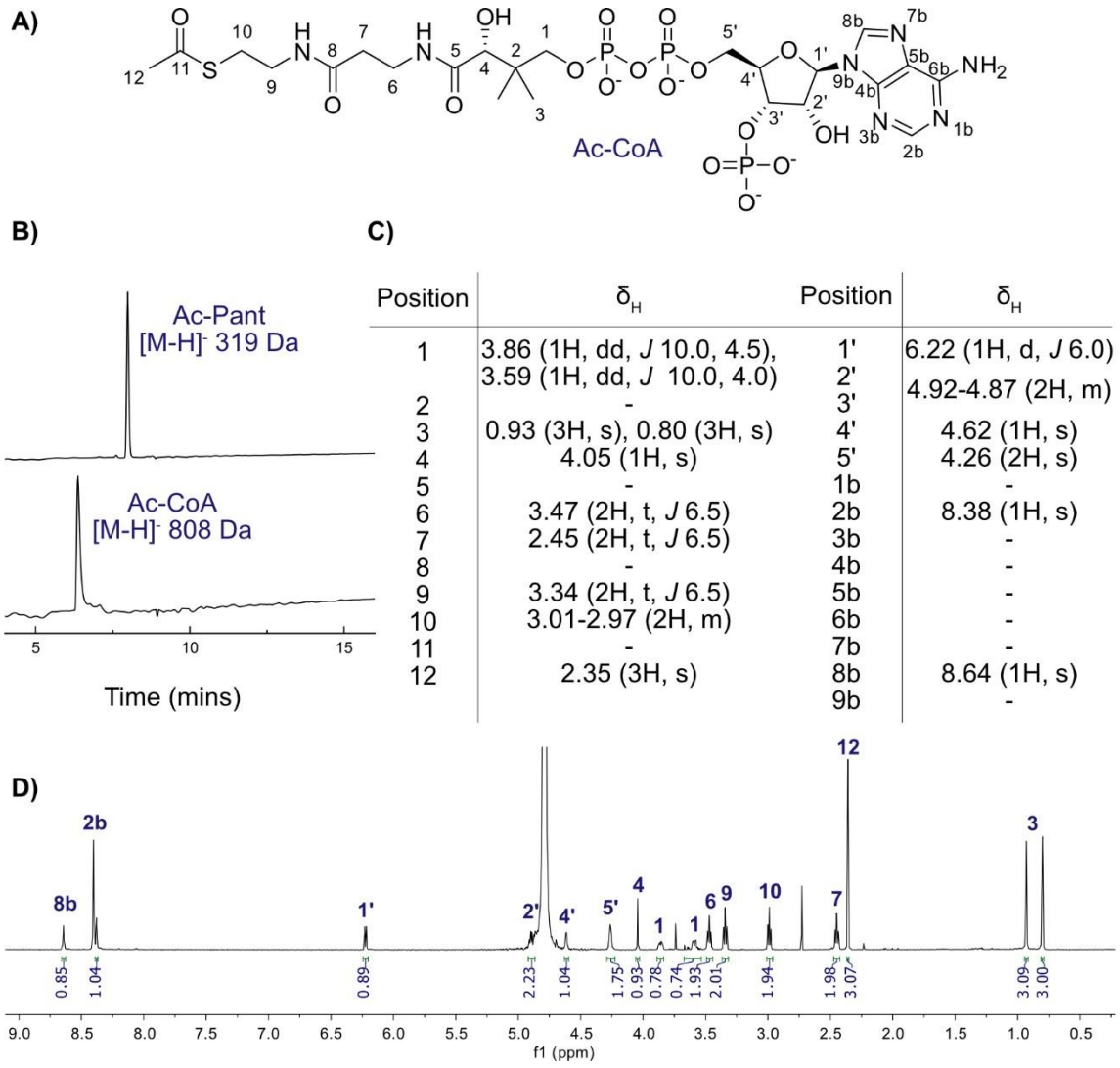


Figure 27 Characterisation of Ac-CoA isolated from the chemo-enzymatic upgrade of Ac-pantetheine. A) The structure of Ac-CoA. B) LC/MS trace showing the full conversion of Ac-pant into Ac-CoA after 1 h. C) Assignment of ^{1}H NMR. D) ^{1}H NMR (D_2O) of Ac-CoA purified by LC/MS.

2.1.5 Mass spectrometry assay for ACP4-6

A key part of preparing protein samples for analysis by MS is removal of salt and small molecules that are more likely to ionise than the protein of interest. This is most commonly achieved through small-scale affinity chromatography where the protein of interest is bound to a column, the salt is washed away and followed by elution into a volatile solvent, for example acetonitrile.

Several desalting methods were trialled during this project. A batch method using free C4 resin was time consuming and resulted in poor recovery of protein and poor ionisation. Dialysis into a volatile buffer (for example, ammonium acetate) was investigated but was time consuming and not suitable for assays involving time-course monitoring. Finally, a

method using C4 resin immobilised in a 10 µL Ziptip (Millepore) was determined to be the optimal method. The resin has a high capacity, the method is fast and good ionisation of the sample was observed.

Having shown that Ac-pantetheine **60** may be converted to Ac-CoA, the one-pot pantetheine upgrade and ACP loading to form Ac-BatA was investigated. Following the literature method of Moore *et al*., apo-BatA, Ac-pantetheine **60** and the CoA upgrade enzymes were incubated with MupN in the presence of $MgCl_2$ and ATP at room temperature for 1 h. The literature method used 9 mM ATP for the conversion and this proved effective in the test reaction. However, after one hour, only 50% ACP loading was observed along with the formation of a fine precipitate. This had not been previously observed for the apo- to holo- conversion of BatA, suggesting that one of the reagents for the CoA upgrade was promoting protein precipitation. The ATP concentration was lowered to 1 mM and following 1 h incubation, a complete conversion of apo-BatA to Ac-BatA was observed as confirmed by MS and Ppant ejection assay (Figure 28).

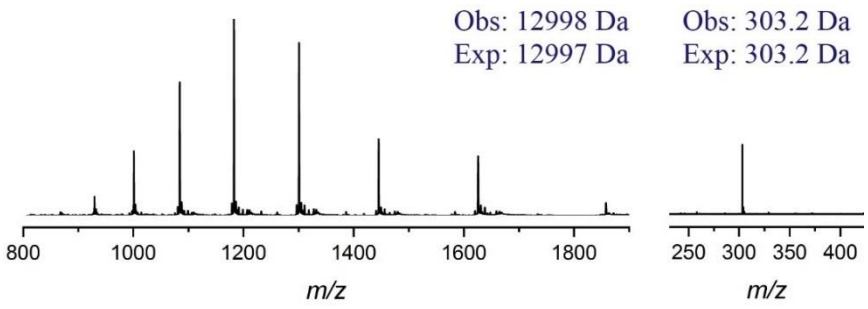
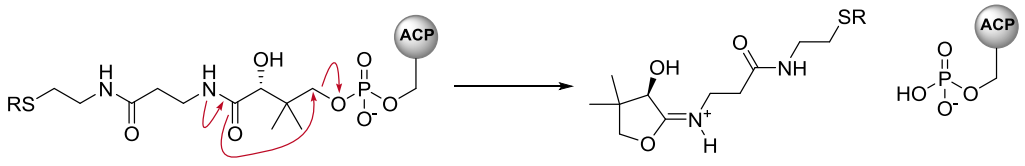


Figure 28 MS analysis of Ac-BatA including the Ppant ejection assay.

The Ppant ejection assay is a MS technique that relies on the selective fragmentation of an ACP. The ions that correspond to a single charge state are isolated and fragmented by collision-induced dissociation (CID), leading to the ejection of a singly charged "ejected" ion and phosphorylated apo-ACP (Scheme 20). Thioester-bound intermediates are stable to the assay conditions, allowing for identification of the tethered species with high mass accuracy.



Scheme 20 Ppant ejection assay occurs by collision-induced dissociation of an ACP, producing a singly-charged Ppant ejection ion.

Following the small-scale proof of concept reaction, large scale conversions of apo-BatA to Ac-BatA were set up in which the concentration of apo-BatA was increased (up to 500 μM) or the reaction volume was increased (up to 10 mL). Longer reaction times, double addition of reagents and agitation of the reaction were sometimes required to get full conversion to Ac-BatA. With careful monitoring by MS, it is possible to convert 10 mg of apo-BatA into Ac-BatA. Once complete, reactions were desalted to remove excess CoA that might interfere with downstream enzymatic assays.

The stability of the acetyl group was monitored by MS and after 24 hours at room temperature, around 50% of the acetyl group had been hydrolysed to holo-BatA. However, when stored at -80 °C, the acetyl group was stable for up to 3 months with negligible hydrolysis to holo-BatA. Therefore, large scale conversions of apo-BatA to Ac-BatA were undertaken and following desalting, the protein was concentrated and aliquoted prior to flash freezing and storage at -80 °C.

Loading of the modular apo-ACPs was achieved by incubation with Acac-pantetheine **62** in an analogous one-pot procedure. Although not quantified, each of the modular ACPs appeared to load at a different rate, but by monitoring the reaction by MS the progress of loading could be followed until completion. If appropriate, reactions would be left longer or more reagents added to the reaction. In particular, the addition of an extra equivalent of MupN would often expedite the reaction to completion. It should be noted that a small amount of holo-ACP was almost always observed during the loading of modular ACPs with β-ketothioester-pantetheines, arising from the hydrolysis of the thioester bond.

The stability of all the modular Acac-ACPs was monitored by MS and the thioester was shown to be more labile than Ac-BatA. After 16 h at room temperature, complete hydrolysis of Acac-ACP to holo-ACP was observed. Additionally, protein degradation was also observed over this time frame. Storage at -20 °C slowed the rate of hydrolysis, but even with this safeguard, significant hydrolysis had occurred after 24 h. PropAc-pantetheine **63** was also prepared and it was hoped that the longer carbon chain would provide some additional hydrolytic stability due to the increased alkyl chain.[133] However, Acac-ACP and PropAc-ACP appeared to be equally labile and both substrates were used interchangeably throughout the course of this work. Due to the lability of the β-ketothioesters, modular ACPs were loaded prior to β-branching assay and used immediately.

During β-branching assays, approximately equimolar concentrations of Ac-BatA and Acac-ACP was desired. Stocks of both ACPs were prepared at around 100 µM and following equal volume mixing and addition of catalytic domains, a final concentration of 50 µM was commonly achieved.

2.1.6 Unsaturated β-methyl branches

Equimolar Ac-BatA and Acac-ACP4 (50 µM) were incubated with BatC (5 µM) at room temperature for 1 h. A sample was desalted and analysed by MS and it was immediately apparent that despite equimolar concentrations, BatA ionised far more efficiently under the MS conditions and approximately 10 times more signal was observed for BatA than ACP4. These signals indicated that there was a loss of Ac-BatA and formation of holo-BatA consistent with the transfer of the acetyl group to BatC, although partial hydrolysis could not be ruled out. Further analysis of low intensity peaks resulted in the identification of a single charge state corresponding to the following ACP4 species: holo-, Acac- and HMG-ACP4 (Figure 29A). The mass shift from Acac-ACP4 to HMG-ACP4 was +60 Da and this was confirmed by Ppant ejection of both species (Figure 29B and 29C). This initial result showed that the acetyl group had been successfully transferred from Ac-BatA to Acac-ACP4 *via* BatC. However, the reaction efficiency was poor with a significant amount of Acac-ACP4 starting material still present as well as holo-ACP4 arising from hydrolysis of the Acac- or HMG-ACP4 thioesters.

BatD and BatE were added to the assay to catalyse the loss of H_2O and CO_2 respectively and form an unsaturated β-branched product. Addition of BatD alone did not result in a mass shift consistent with the loss of water. This was to be expected as the BatD-catalysed dehydration is a reversible reaction and so the observation of product would depend on the equilibrium position. A small amount of dehydrated species has been characterised before by Calderone and co-workers,[39] however, due to the difficulty in observing this species, most *in vitro* studies on the HCS cassette couple the dehydration and decarboxylation steps.[40] This increases the complexity and uncertainty in the assays which may yield no results or unexpected species, due to the two proteins acting with only a single analytical point.

Addition of both BatD and BatE consumed HMG-ACP4, but due to the 2 Da mass difference between Acac-ACP4 and enoyl-ACP4 overlap of the charge states at 1592 Da was observed (Figure 29A). Ppant ejection resulted in the observation of two ions,

corresponding to Acac-ACP4 at 359.2 Da and enoyl-ACP4 at 357.2 Da (Figure 29D). This was the first successful reconstitution of the β-branching pathway from the kalimantacin gene cluster, however, the efficiency of the transformation was generally poor.

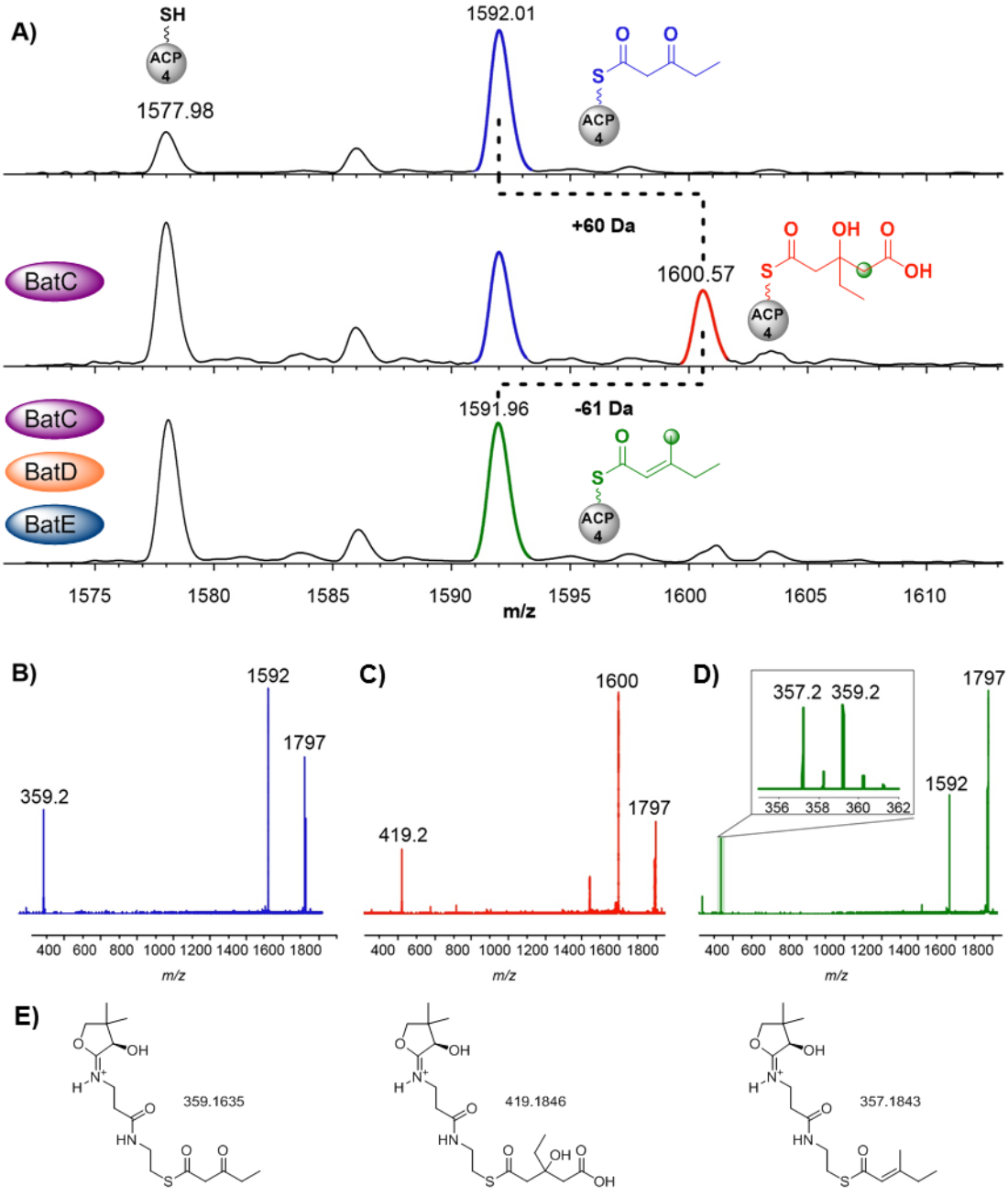


Figure 29 β-branching MS assay for ACP4 by monitoring the 7^+ charge state. A) Single charge state analysis of PropAc-ACP4 (top, 1592 Da) being converted into HMG-ACP4 (middle, 1600 Da) by BatC and the subsequent conversion to enoyl-ACP4 (bottom, 1592 Da) by treatment with BatD/BatE. The charge state at 1578 Da corresponds to holo-ACP4, arising from acyl-ACP4 hydrolysis. B) Ppant ejection from PropAc-ACP4 at 1592 Da gives a 1^+ ion at 359.2 Da and a 6^+ ion (phosphorylated ACP4) at 1797 Da. C) Ppant ejection of HMG-ACP4 at 1600 Da gave a 1^+ species at 419.2 Da. D) Ppant ejection of 1592 Da with zoomed area separating the 1^+ ions of unreacted PropAc-ACP4 (357.2 Da) and enoyl-ACP4 (359.2 Da). E) The structure of the Ppant ejection ions and their exact masses.

Optimisation of the condensation step based around the previous literature reports using TCEP, shorter reaction times and varying the equivalents of reagents to increase our yield of HMG-ACP4 did not result in a dramatic change in the ratio of species.[39,41,45] A typical reaction profile exhibited approximately $^{1}/_{3}$ holo-ACP4, and varying ratios of Acac-and HMG-ACP4. The best results achieved were $^{1}/_{3}$ holo-ACP4, only trace amounts of Acac-ACP4 and ~ $^{2}/_{3}$ overall conversion to HMG-ACP4. No difference in outcome was observed when using frozen aliquots of BatC compared to freshly purified protein. Similarly, the BatD/BatE coupled step was also optimised with the best results achieved using freshly purified protein that was stored at 4 °C for less than 5 days. Inconsistent results were obtained when using frozen aliquots or storage at 4 °C for longer than 1 week. As a result, these assays often required a higher catalyst concentration, double addition of reagents or longer reaction times.

Due to the abundance of β-branching ACPs in the kalimantacin gene cluster and our interest in understanding the formation of the three different, consecutive β-methyl branches, the assay was also reconstituted with ACP5 and ACP6 (Figure 30).

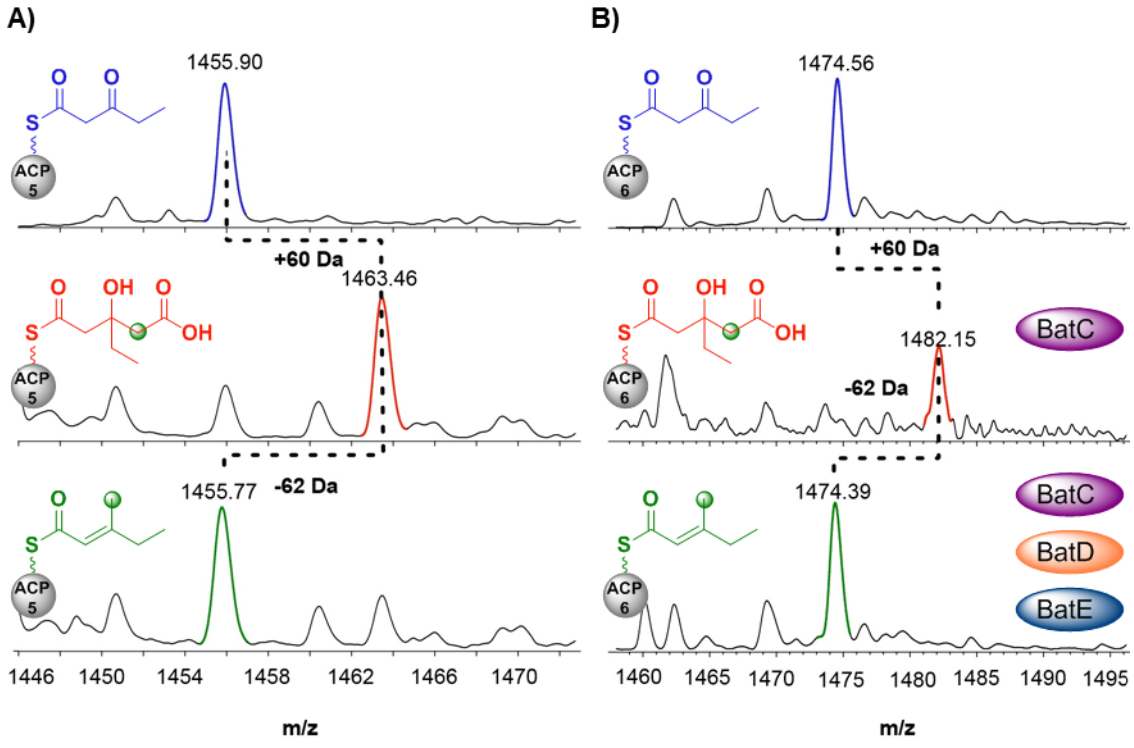


Figure 30 Single charge state (8^{+}) monitoring for the β-branching assay for ACP5 (left) and ACP6 (right). PropAc-ACP (top) is converted to HMG-ACP (middle) when reacted with Ac-BatA and BatC, followed by BatD/BatE-catalysed conversion to enoyl-ACP (bottom).

Once again, both ACPs showed a lower ionisation efficiency under the MS conditions than BatA, adding to the difficulty of analysis. Both ACPs successfully formed the HMG-ACP intermediate by transfer of acetyl from Ac-BatA to PropAc-ACP by BatC. This was subsequently converted to into enoyl-ACP by addition of BatD and BatE. Less hydrolysis and a cleaner conversion of PropAc-ACP to enoyl-ACP was observed for ACP5 and ACP6. Unfortunately, a subsequent computer failure meant a significant loss of MS data was encountered. As such, not all Ppant ejection assay data or whole spectra are presented.

In vitro β-branching assays monitored by LC/MS

MS analysis of these ACPs proved to be a valuable tool in initially characterising the pathway. However, a variety of analytical methods were investigated for monitoring the challenging *in vitro* reconstitution of this multi-enzyme pathway. LC/MS analysis provides a clearer detection of each species due to separation of the highly ionisable BatA and the separation of different modular ACP-bound intermediates.

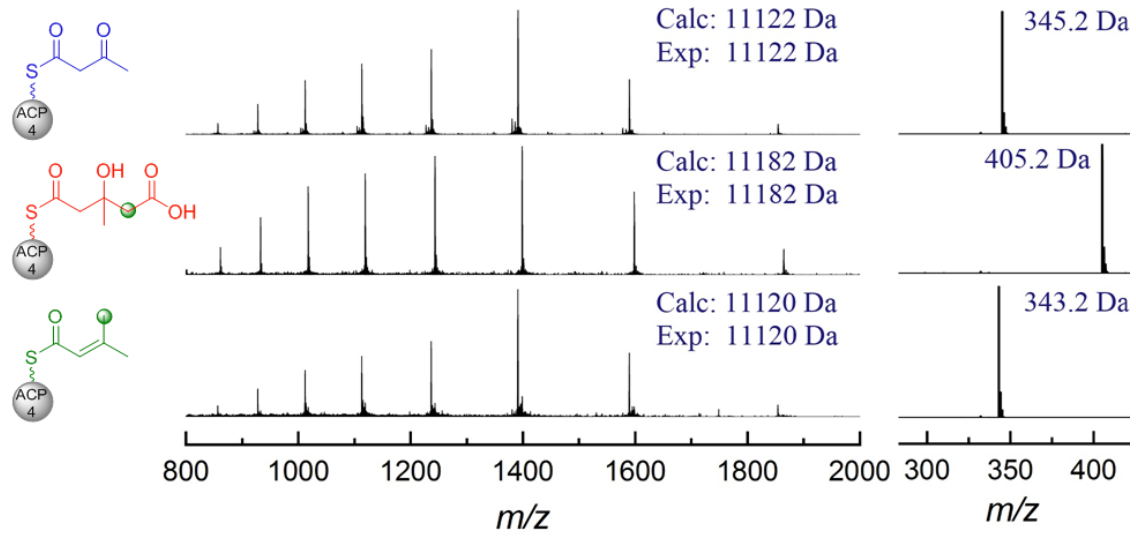


Figure 31 LC/MS analysis of β-branching on ACP4. A full envelope of charge states and Ppant ejection ions are seen for Acac-ACP5 (blue), HMG-ACP5 (red) and enoyl-ACP5 (green).

Injection of Acac-ACP4 alone gave a full envelope of charge states with good mass accuracy when deconvoluted (Figure 31). Ejection of the most abundant charge state gave a 1^+ ion at 345.2 Da corresponding to Acac-ACP4. Transfer of acetyl from Ac-BatA to Acac-ACP4 *via* BatC resulted in the formation of HMG-ACP4, confirmed by mass shift of the envelope of charge states and Ppant ejection (405.2 Da). Ac/holo-BatA were eluted

off the column prior to the ACP4-bound intermediates and did not interfere with ACP4 analysis. HMG-ACP4 eluted at an earlier retention time than Acac-ACP4 resulting in a single envelope of charge states. Addition of BatD and BatE to the assay catalysed the conversion of HMG-ACP4 to enoyl-ACP4. Enoyl-ACP4 was the last of the ACP4-bound species to be eluted from the column and once again this species was characterised by a full envelope of charge states and a Ppant ejection ion at 343.2 Da.

To determine whether separation could be achieved for other ACPs, the assay was undertaken with ACP5 and the data presented in Figure 32. Once again, Ac/holo-BatA eluted prior to the ACP5-bound intermediates, which eluted in the same order as observed for ACP4: HMG-, Acac- and enoyl-ACP5. A full envelope of charge states and the complementary Ppant ejection data was collected and confirmed the formation of an unsaturated β-methyl branch.

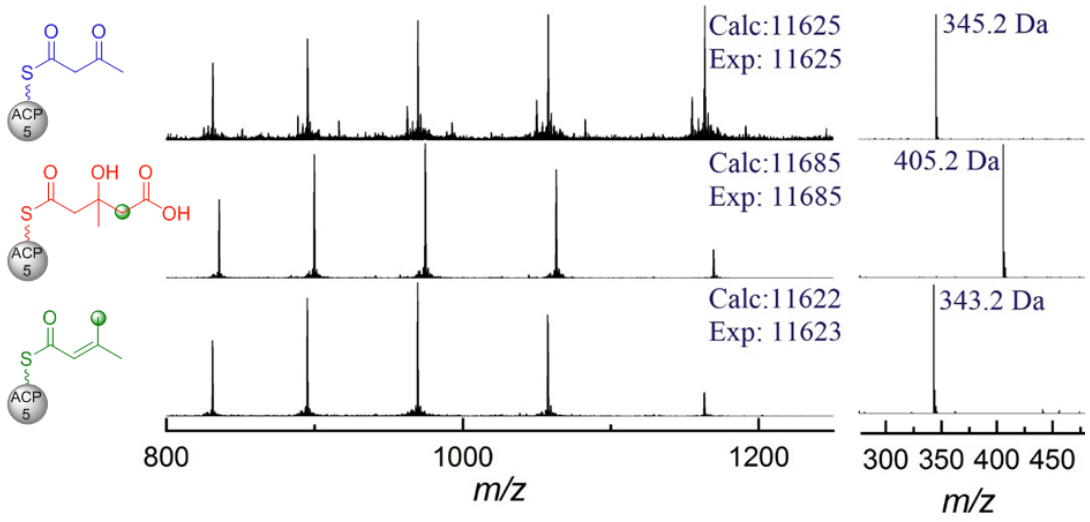
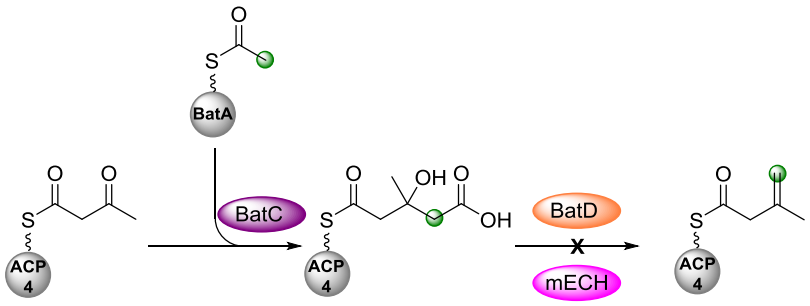


Figure 32 LC/MS analysis of β-branching on ACP5. A full envelope of charge states and Ppant ejection ions are seen for Acac-ACP5 (blue), HMG-ACP5 (red) and enoyl-ACP5 (green).

With the successful MS assay in hand, attempts were made to determine the role of mECH in the formation of β-branches. Direct detection of either an *endo*-β-methyl or *exo*-β-methyl branch cannot be made by MS due to the identical masses of both species. However, its broad role can be inferred using MS if a domain of known function can be complemented by the mECH domain. It was therefore decided to replace BatE with mECH to determine whether the modular domain can act with a similar decarboxylase activity. Using the LC/MS technique to monitor the assay, Acac-ACP4 was first reacted with Ac-BatA and BatC to form HMG-ACP4. BatD and mECH were both added to the

reaction mixture and allowed to incubate at room temperature. The reaction was monitored (up to 6 h) by LC/MS, however, no peak corresponding to an unsaturated β-methyl was observed. Heating the reaction did result in a trace amount of species corresponding to an unsaturated β-methyl branch, however, this species was also observed in the negative control reaction, containing only HMG-ACP4, and arises from thermal decarboxylation. These results suggested that mECH cannot complement for BatE (Scheme 21).



Scheme 21 Unsuccessful complementation of BatE by mECH.

2.1.7 Saturated-β-methyl branches

In addition to the two unsaturated β-methyl branches formed in the biosynthesis of kalimantacin, two sat.-β-methyl branches are also formed. Firstly, when the biosynthetic intermediate is attached to module 6 located on Bat2, the second large ORF in the gene cluster, and secondly, when the intermediate is attached to Bat3 ACP5 (module 12) in the middle of a series of three consecutive β-methyl branches. The sat.-β-methyl is hypothesised to be formed *via* the reduction of an HCS cassette derived *endo*-β-methyl by a *trans*-acting ER domain. BatK is the only ER in the gene cluster and is a free-standing domain. Lavigne *et al.* used yeast two-hybrid (Y2H) to probe interacting proteins within the kalimantacin assembly line and they demonstrated that BatK had a specific interaction with Bat3 ACP2 and ACP5.[134] These interactions are consistent with the proposed biosynthetic pathway whereby an unbranched α,β-unsaturated moiety is reduced at ACP2 and at ACP5 to form the sat.-β-methyl.

Having shown the formation of the *endo*-β-methyl on ACP5, the ER-catalysed reduction of a β-branched intermediate was then investigated (Figure 33). ACP5 was loaded with enoyl-pantetheine **64** under standard one-pot conditions. BatK and NADH were added and the reaction monitored by MS, which showed complete consumption of the starting material (343 Da) by Ppant ejection assay and the formation of a new Ppant ejected ion

(345 Da) after 1 h. The 2 Da mass increase corresponds to the reduction α,β-unsaturated double bond by addition of H^- from NADH and reprotonation by H^+ (Figure 33B). To test whether the reaction was ACP-specific, the reaction was also performed on ACP4 and ACP6. ACP4 was loaded with enoyl-pantetheine **64**, incubated with BatK and NADH, and after 1 h complete reduction had taken place (Figure 33A). However, enoyl-ACP6 under the same reaction conditions, was not reduced to the sat.-β-methyl (Figure 33C). Leaving the reaction for up to 6 h did not result in reduction of the alkene and no reduction was observed by incubating any of the ACPs with NADH in the absence of BatK.

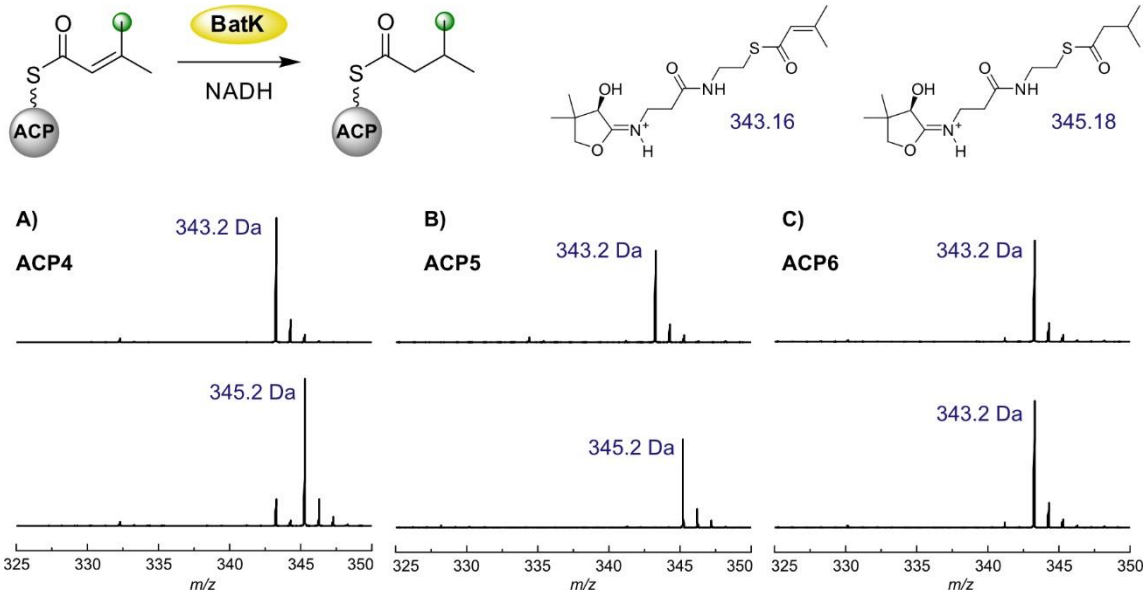


Figure 33 Reduction of enoyl-ACP by BatK and NADH was monitored by the Ppant ejection assay which results in 1^+ ions corresponding to an unsaturated-β-methyl (343.2 Da) or saturated-β-methyl (345.2 Da) intermediate. A) An unexpected reduction by BatK is observed for ACP4. B) Reduction by BatK is observed for ACP5 as expected from the proposed biosynthetic pathway. C) No reduction is observed for the intermediate attached to ACP6. No reduction occurs in the absence of BatK.

These results suggest that protein-protein interactions are only partially significant in the formation of sat.-β-methyl branches. The reduction at ACP5 and the lack of reduction of an intermediate attached to ACP6 were to be expected due to the β-branches incorporated in the native system. However, the interaction of ACP4 with BatK demonstrates its ability to deliver substrates to a wide range of catalytic domains. ACP4 does not need to discriminate against BatK presumably because the *exo*-β-methyl branch is not a substrate of BatK. Remarkably, ACP4 interacts with *trans*-acting domains from the HCS cassette (BatC, D, E), the *trans*-acting AT (BatJ) and now BatK. Additionally, ACP4 must correctly interact and transfer substrates to neighbouring modular domains by bearing a malonyl group that is used to undergo chain extension with the upstream KS of module

10 and, following β-branching, pass the modified chain to the downstream KS of module 11. To achieve correct fidelity through the pathway, control at this ACP must be achieved by reaction kinetics (in *cis* v in *trans*) or steric occlusion from catalytic active sites.

The *in vitro* activity is further supported by the isolation of minor metabolites from the *P. fluorescens* wild-type strain. One full-length and two truncated kalimantacin analogues were isolated, characterised and their antibiotic activity determined by Angus Weir during his Master's research project (Figure 34). Compound **69** is a full-length kalimantacin analogue that possesses a sat.-β-methyl branch at C-7 rather than an *exo*-β-methyl. This likely occurs from aberrant *endo*-β-methyl formation and subsequent reduction by BatK. The two truncated metabolites still possess the *exo*-β-methyl branch but lack either the sat.-β-methyl **70** or *endo*-β-methyl branch **71**, likely due to module skipping. Aside from interest in their biosynthetic origin, the MIC values show that **69** suffers from an approximately 100-fold loss in antibiotic activity due to the C-7 reduction and the truncated intermediates **70**-**71** lose all antibiotic activity. This intriguing structure-activity relationship (SAR) has been followed-up by Angus Weir during his PhD studies.

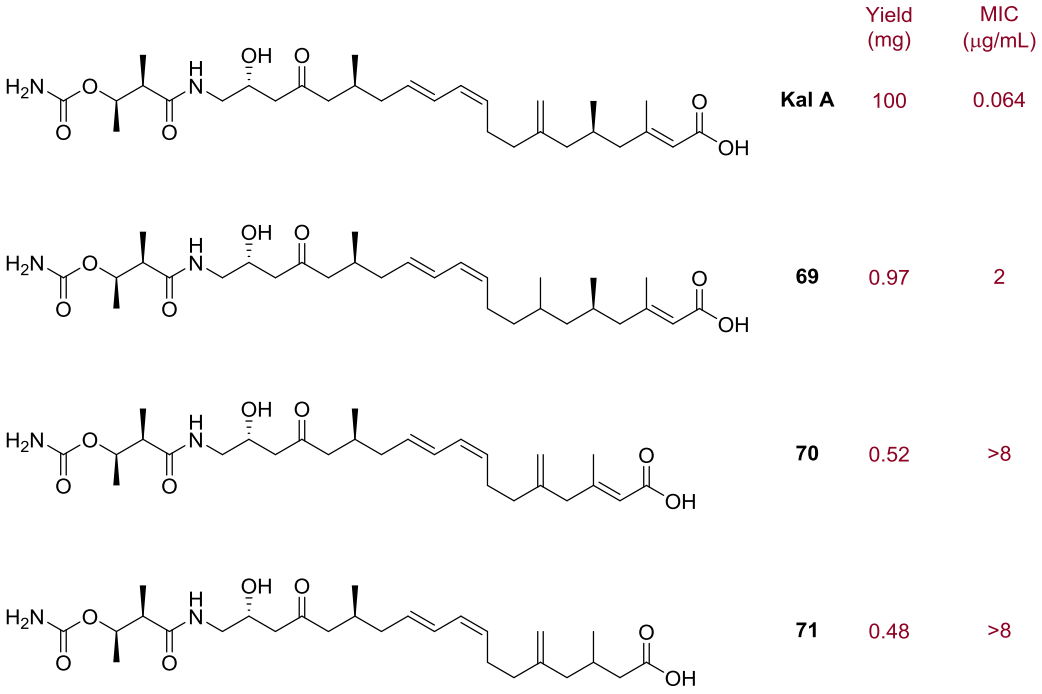


Figure 34 Minor metabolites **69**-**71** isolated from *P. fluorescens* WT strain and their yields and MIC values are compared against kalimantacin A.

Considering both the *in vitro* observations and the minor metabolites isolated, ACP4 does achieve the remarkable selectivity required to produce kalimantacin as the major isolate. Whilst some clarity has been provided for the formation of the three different β-branches in kalimantacin, there are still unknown answers in relation to the selectivity achieved.

To conclude, the results presented demonstrate the successful *in vitro* reconstitution and characterisation by MS of the HCS cassette from the kalimantacin pathway. The complexity of the β-branching pathway required the preparation of many biosynthetic intermediates and proteins, as well as numerous analytical techniques to probe the outcome. All *trans*-acting domains (BatA, C-E) were functionally active and each of the consecutive ACPs (ACP4-6) were able to interact with the cassette to produce an unsaturated β-branched product. The geometry of the unsaturated product was not determined, although the working hypothesis made it likely that an *endo*-β-methyl was produced. The ability of BatK to reduce an *endo*-β-methyl branch on ACP4 and 5 was proven, as well as the lack of reactivity with ACP6. Finally, a previously unidentified modular ECH domain has been identified within Bat3. In this initial phase of the work, the domain was expressed as a soluble protein, however, no function has been shown so far. This lack of function may point to a mis-folded, non-functional protein that requires neighbouring domains or correct oligomerisation for activity.

2.2 NMR Characterisation of β-Branching

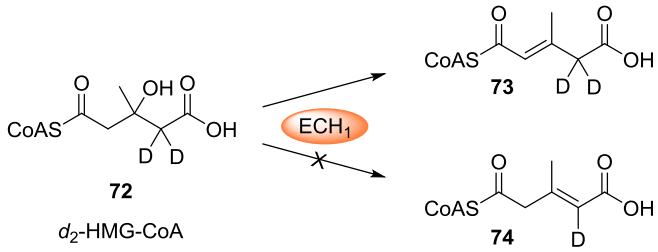
2.2.1 Pantetheine NMR experiments

Mass spectrometry, in particular the Ppant ejection assay, has been a powerful tool for the analysis of the β-branching pathway.[39,45] Small sample volumes, dilute assay conditions and high mass accuracy have enabled careful dissection of each catalytic step. However, kalimantacin possesses three types of β-methyl branch, two of which are indistinguishable by mass alone. Previous *in vitro* characterisation of alkene β-branches has relied on either UV spectroscopy assaying for the α,β-unsaturated thioester chromophore or hydrolysis of ACP-bound intermediates and subsequent small molecule analysis by GC-MS.[39,43,45] It was felt that both of these methods were less than robust and an improved analytical technique was sought.

So far, it has been assumed that the product of reconstituting the HCS from the kalimantacin pathway is an *endo*-β-methyl as it is the most common β-branch observed in polyketides and our hypothesis suggests that other β-branches may be formed from this

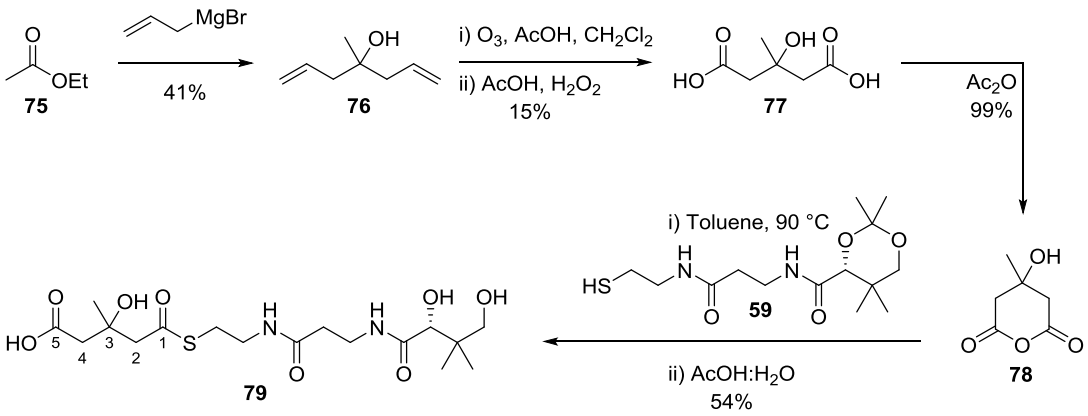
intermediate. Conducting β-branching assays and monitoring by NMR would allow for the real time monitoring of such reactions *in situ*. This could be undertaken with a small molecule mimic of the β-branching intermediates to avoid the use of stoichiometric ACPs which would dominate the spectrum.

Since completing the work described in this thesis, Smith *et al.* reported the ECH_1- catalysed α,β-dehydration of HMG-CoA, despite the β-branching product containing a non-decarboxylated β,γ-unsaturated acrylate moiety.[63] This was achieved by deuterium labelling to produce d_2-HMG-CoA **72** followed by NMR and MS analysis confirming the presence of the dehydrated product **73** and no evidence of β,γ-dehydration product **74** (Scheme 22). Proton exchange with the solvent, the kinetic isotope effect and low turnover are significant limitations to this methodology.



Scheme 22 Formation of an α,β-unsaturated dehydration product was confirmed by MS and NMR.

It was decided to synthesise HMG-pantetheine **79** and monitor its reaction with BatD and BatE. There is literature precedent for the synthesis of the key anhydride intermediate **78**, which has been previously coupled to SNAC. However, SNAC derivatives have been shown to be inactive with ECH_1/ECH_2 homologs from the bacillaene pathway.[135] Therefore HMG-pantetheine **79** was chosen as the synthetic target as it represented a superior biomimetic. The initial aim was to use the small molecule for NMR studies, although, the pantetheine allows for the loading onto an ACP if required. A disadvantage of the literature synthetic route is the non-stereoselective formation of HMG-pantetheine which will result in diastereomers at the C-3 position of **79**. Based on literature reports, *S*-HMG is the product of HMGS condensation, matching the stereochemical outcome of HMGCS from primary metabolism, and so a maximum 50% turnover would be expected from the final synthetic pantetheine.[40,41] However, this was not deemed to be a problem as the other diastereomer should not interfere with the assay and HMG-pantetheine **79** was synthesised (Scheme 23).



Scheme 23 Synthesis of HMG-pantetheine **79**.

Adapting the literature procedure,[36] freshly-distilled ethyl acetate **75** was reacted with allylmagnesium bromide to give alcohol **76**, which was treated with ozone followed by an acidic workup. The crude reaction mixture was not concentrated under reduced pressure in case of incomplete degradation of the ozonide, instead saturated aqueous sodium sulfite was added to quench any remaining peroxide and following extraction into ethyl acetate the required diacid **77** was isolated in 15% yield. To maximise recovery, the combined aqueous phases were concentrated to dryness, treated with acetic anhydride and following work-up anhydride **78** was isolated in a low yield, but enough material was recovered to complete the route. Pure di-acid **77** was also converted to anhydride **78** in quantitative yield. Anhydride **78** was ring-opened with protected pantetheine **59** by heating in toluene, and final acetal deprotection under acidic conditions afforded HMG-pantetheine **79**.

HMG-pantetheine **79** was dissolved in NMR buffer (50 mM Na phosphate, 100 mM NaCl, 10% D_2O, pH 8.0) and a control ^1H NMR spectrum was recorded (Figure 35). BatD and BatE were then added to the sample and the ^1H spectrum was recorded at regular intervals. After 45 minutes a small peak was visible at 6.18 ppm. The intensity of the peak increased with time and after 13 h the peak had developed an upfield shoulder (6.17 ppm). This grew further and after 150 h, the original peak had disappeared leaving just a single signal at 6.17 ppm.

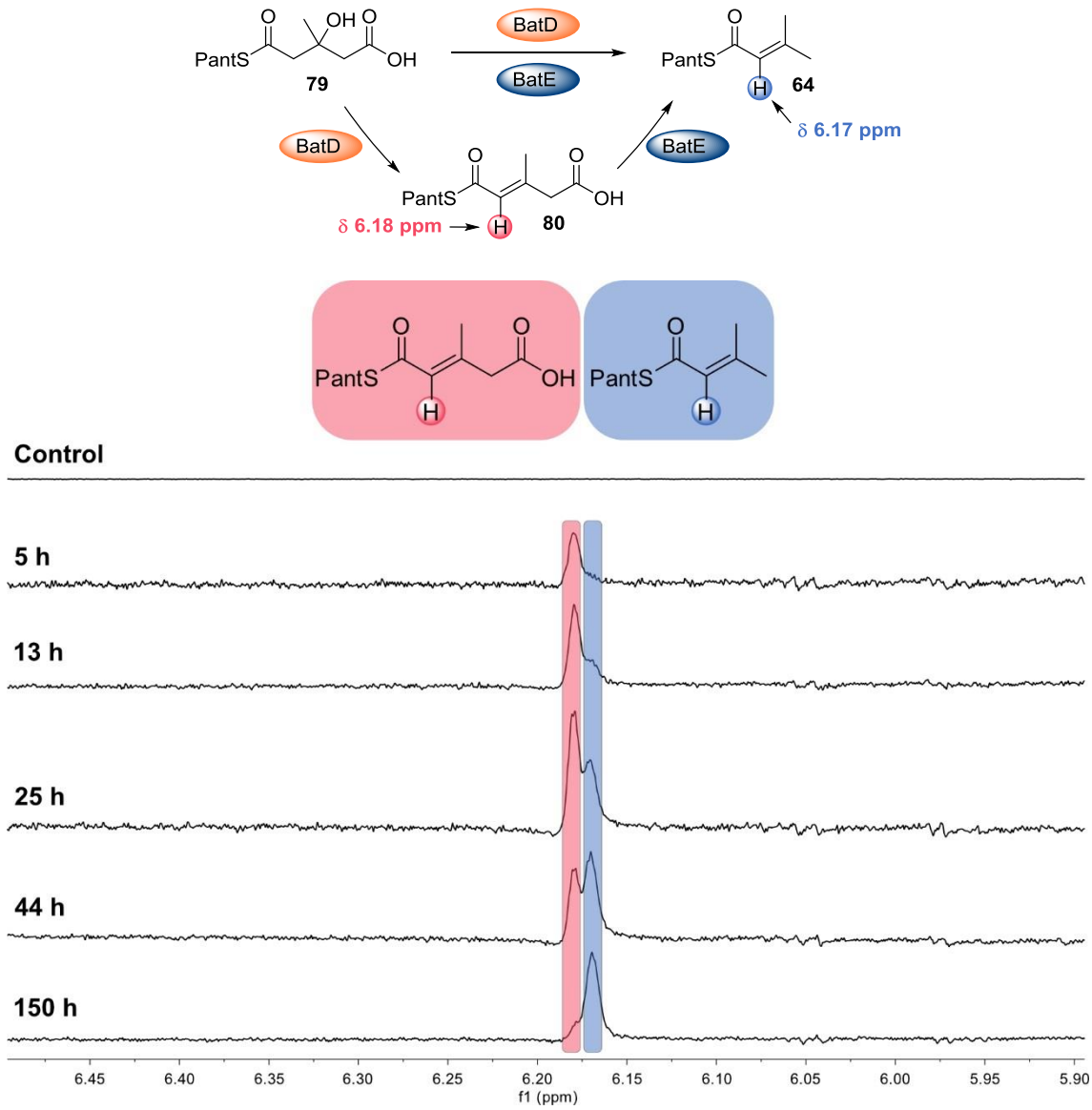


Figure 35 Stacked plot of ^{1}H NMR spectra showing the slow conversion of HMG-pantetheine **79** to enoyl-pantetheine **64** catalysed by BatD and BatE. The peak at 6.18 ppm (red) corresponds to the alkene proton of MG-pantetheine **80** and the peak at 6.17 ppm the alkene proton of enoyl-pantetheine **64**.

A two-dimensional TOCSY experiment confirmed the identity of the two peaks (Figure 36). The downfield peak (6.18 ppm) was assigned to the alkene proton of MG-pantetheine **80** and gave clear correlations to a CH_2 at 3.07 ppm and a CH_3 at 2.14 ppm. The upfield shifted peak (6.17 ppm) corresponded to the alkene peak of enoyl-pantetheine **64** with correlations to two CH_3 groups at 1.91 and 2.11 ppm. The spectra and correlations observed are consistent with those of the synthetic pantetheine **64**.

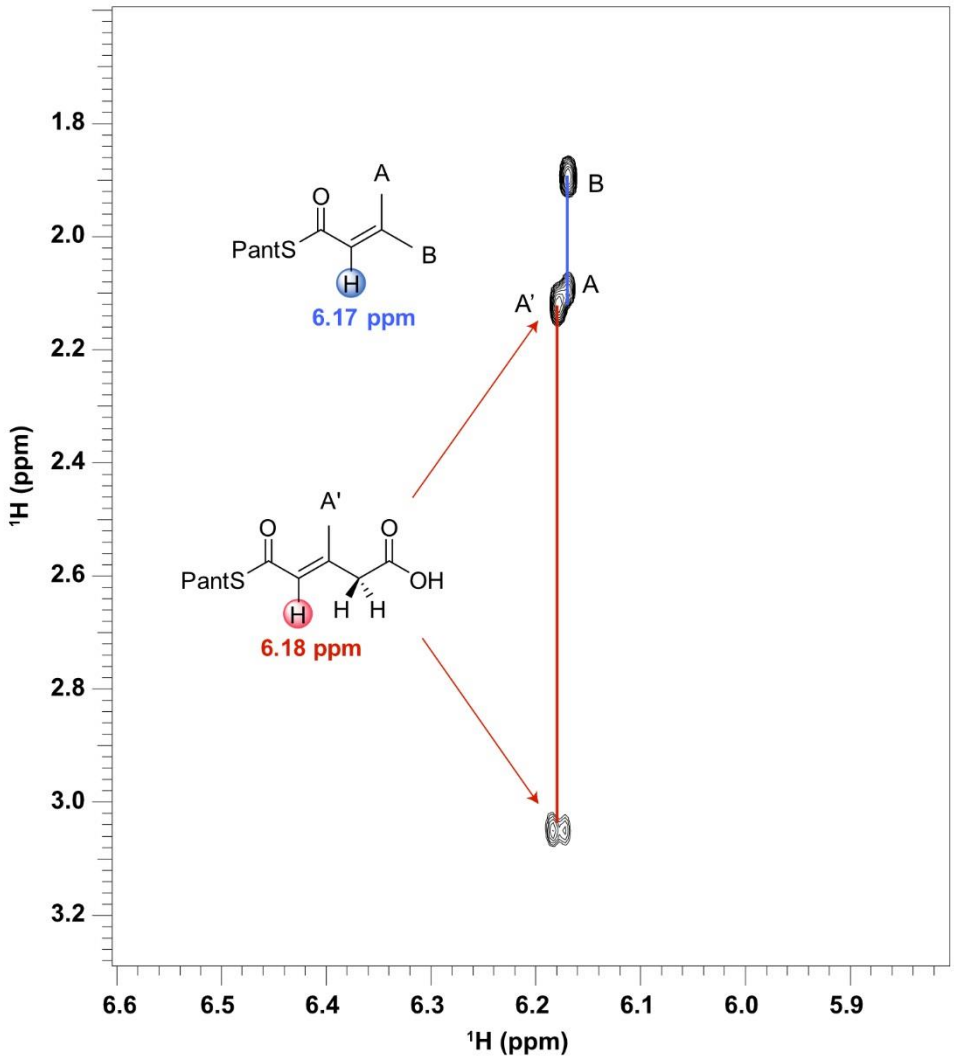
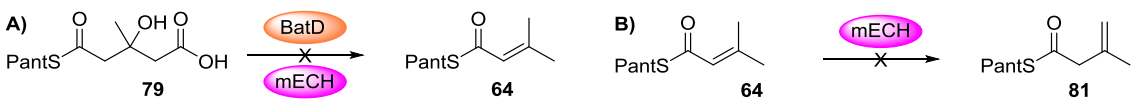


Figure 36 Two-dimensional TOCSY spectrum showing the correlations from the alkene protons of the MG-pantetheine **80** (6.18 ppm) and enoyl-pantetheine **64** (6.17 ppm).

Although mECH showed no activity as a decarboxylase by MS, we sought to validate this result by NMR. HMG-pantetheine **79** was incubated with BatD and mECH and the build-up of MG-pantetheine **80** was observed. Further monitoring over time did not produce any evidence for a product of decarboxylation (Scheme 24A). Finally, enoyl-pantetheine **64** was incubated with mECH to monitor for isomerism of the α,β-unsaturated bond, however, no loss of signal or formation of a new peak corresponding to an *exo*-β-methyl **81** was observed (Scheme 24B).



Scheme 24 Unsuccessful reactions with mECH by NMR. A) mECH did not complement for BatE as a decarboxylase. B) No isomerism activity was observed in the conversion of an *endo*-β-methyl to an *exo*-β-methyl.

A two-dimensional TOCSY spectrum of MG-pantetheine **80** showed correlations to a CH_2 at 3.07 ppm and CH_3 at 2.14 ppm. To fully characterise MG-pantetheine **80**, a two-dimensional NOESY spectrum was then used to assign the geometry of the alkene (Figure 37). A correlation was seen from the alkene proton only to the CH_2 protons at 3.07 ppm, with no correlation to the CH_3 at 2.14 ppm. This confirmed the syn-assignment of the CH_2 and alkene proton, and the arrangement of the methyl group (that would normally represent the polyketide chain) *trans* to the proton (*E*-isomer).

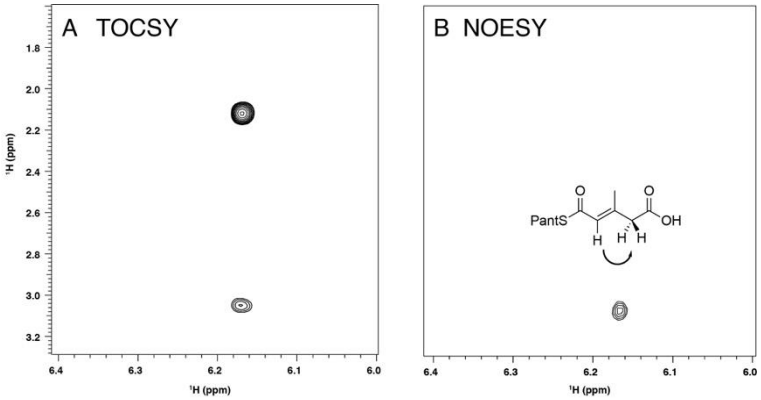


Figure 37 Two-dimensional NMR assignment of MG-pantetheine **80**. A) TOCSY correlations from the alkene proton at 6.18 ppm to the CH_3 at 2.14 ppm and a CH_2 at 3.07 ppm. B) NOESY spectrum showing a correlation from the alkene proton at 6.18 ppm only to the CH_2 at 3.07 ppm, confirming the *cis*-arrangement (*E*-alkene).

These experiments confirmed the formation of an *endo*-β-methyl branch from BatD-catalysed dehydration and BatE-catalysed decarboxylation of HMG-pantetheine **79**. Additionally, the *in-situ* accumulation of MG-pantetheine **80** confirmed the *E*-geometry of the trisubstituted double bond, however, no evidence of mECH activity was observed.

The slow rates of the reaction observed may be due to poor selectivity for a pantetheine substrate mimic, rather than an ACP-bound intermediate. Therefore, developing an assay that allowed for monitoring of ACP-bound intermediates by NMR was required. Such a method exists in the literature and has been used within our group previously.[136,137] It relies on the uniform incorporation of ^{13}C and ^{15}N into the ACP backbone and the loading of a natural abundance Ppant arm onto the ACP. NMR isotope filtration experiments can then be recorded whereby all protons attached to either ^{13}C or ^{15}N are filtered (not observed), allowing for the recording of a ^{1}H spectrum that only contains signals from the Ppant arm and any thioester-bound substrates.

A $^{13}C/^{15}N$ double-labelled sample of ACP4 was prepared by growth in minimal medium supplemented with ^{13}C glucose and $^{15}NH_4Cl$ as the sources of isotopic labels. The ACP was purified as per the unlabelled material, loaded with enoyl-pantetheine **64** and a ^{1}H spectrum of the protein was recorded (Figure 38). Upfield shifted methyl peaks that are representative of a folded protein were visible, and they were split into doublets resulting from spin-coupling with the enriched ^{13}C.

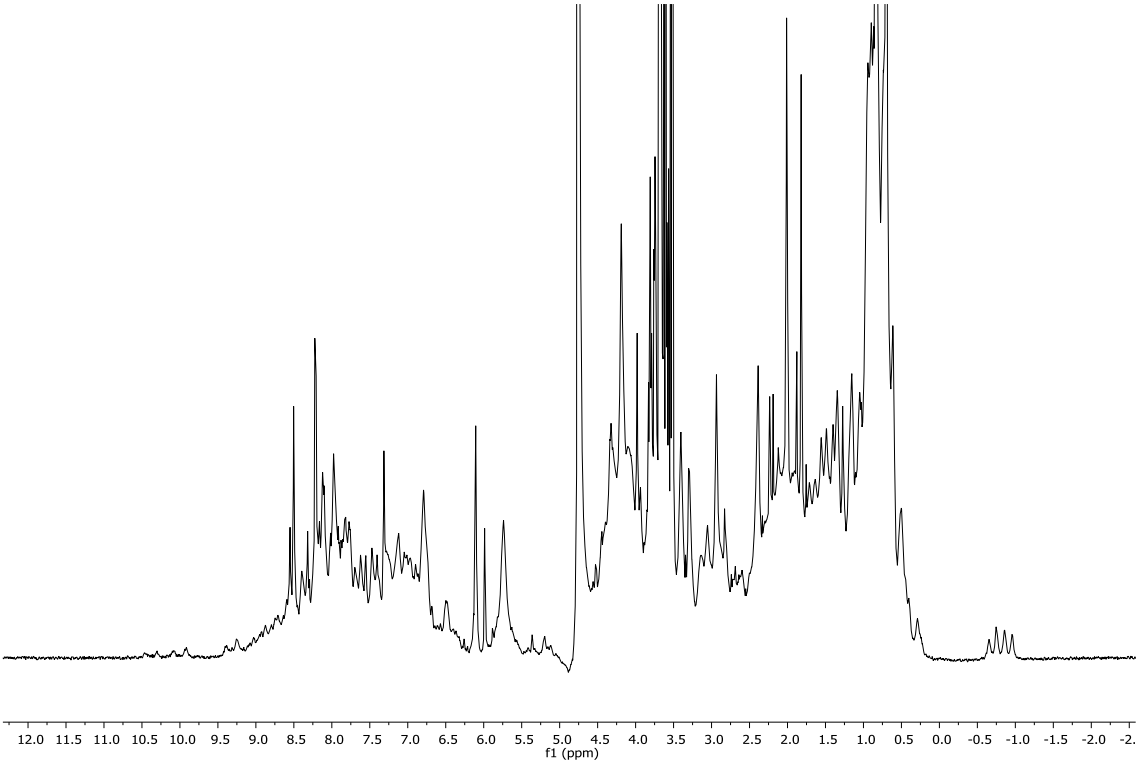


Figure 38 ^{1}H NMR spectrum of enoyl-$^{13}C/^{15}N$-ACP4. The upfield shifted methyl peaks between -0.5 and -1.0 indicate the protein is folded.

A control filtration experiment was recorded that showed the characteristic peaks for a Ppant arm as well as the methyl signals for the unsaturated thioester. However, incomplete labelling meant the spectrum also contained unlabelled protein signals (Figure 39). mECH was added to the sample and it was monitored by ^{1}H NMR. No loss of intensity of the methyl peaks or new peaks were observed, even after the second addition of mECH domain, suggesting that either mECH was not functional or it does not act as an isomerase. The reproducibility of the spectrum over 8 hours was testament to the stability of the ACP-bound intermediate. An attempt was made to load HMG-pant onto ACP4 to monitor the BatD/BatE-catalysed conversion to an *endo*-β-methyl. However, the spectrum was too crowded to assign unambiguously and monitor due to unlabelled protein signals.

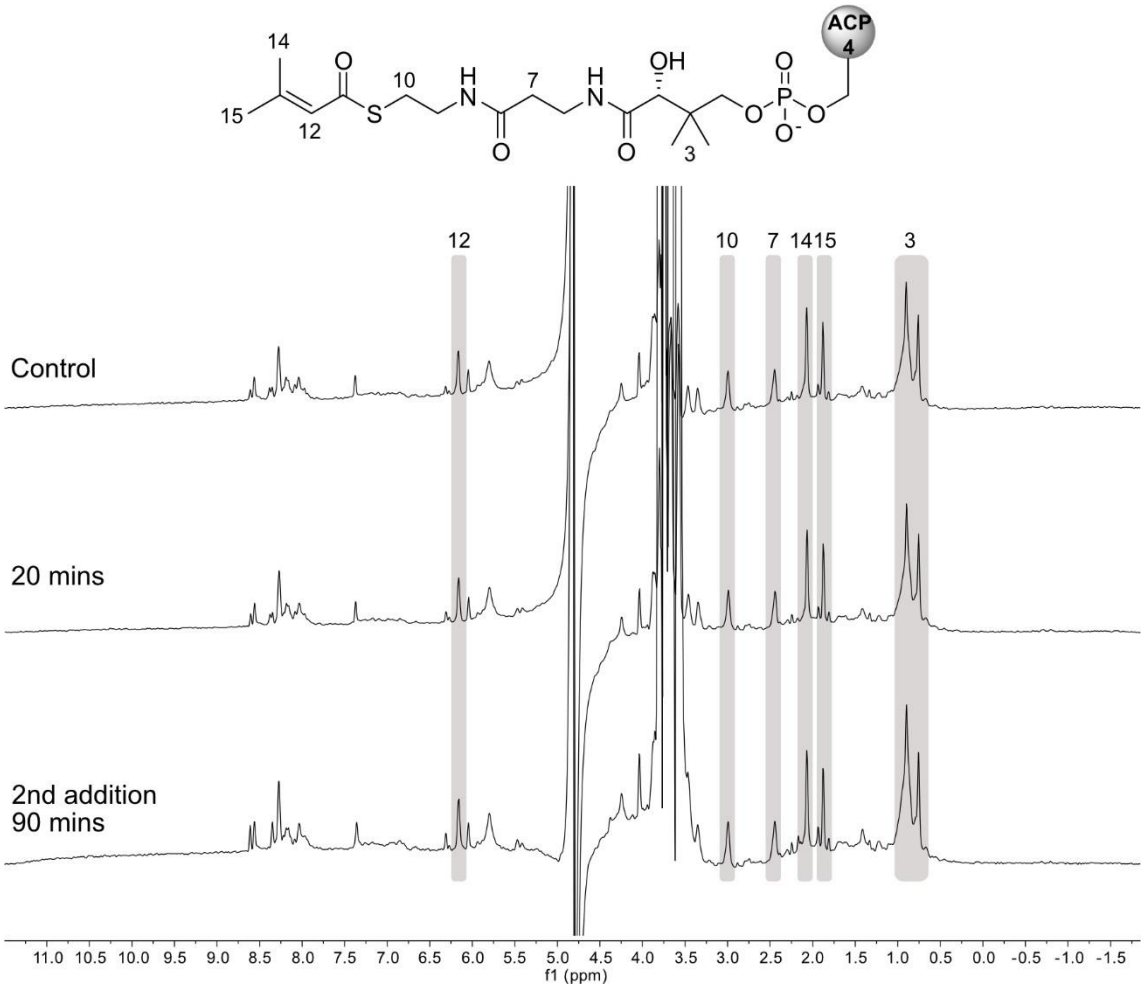


Figure 39 Stack of $^{13}C,^{15}N$-filtered NOESY spectra for enoyl-$^{13}C/^{15}N$-ACP4 + mECH. Signals corresponding to the Ppant arm were observable (grey), however, additional protein peaks arising from incomplete [^{13}C]/[^{15}N] labelling made identifying new peaks challenging.

The filtration experiments were a useful tool for monitoring the potential isomerism reaction, due to the sharp methyl peaks in a clear region of the spectrum. The major limitation of the methodology is the production of uniformly labelled protein. Growth in minimal media alone often resulted in cell death or minimal expression, whilst only 90% labelling was obtained when LB media was used for an initial growth prior to media exchange into minimal media and expression. As a result, the technique was not developed further.

The conclusion from the MS and NMR work undertaken so far was that the mECH domain was not acting as expected. It did not show activity to complement BatE by MS and NMR or show any signs of isomerising *endo*-β-methyl branches by NMR. A literature review revealed that Smith *et al.* were unable to crystallise JamJ, the ECH_2 domain from the jamaicamide biosynthetic pathway that results in the formation of an *exo*-β-methyl.[55]

Close inspection of the domain boundaries reveals that JamJ was cloned without the C-terminal helix found in CurF and present in other ECH_2 domains including BatE.[40] When designing a construct to clone mECH from Bat3, secondary structure prediction and homologous structures suggested inclusion of this helix. However, when excising a domain from a modular PKS, the selection of boundaries is critically important. As such, site-directed mutagenesis was used to create a truncated mECH domain (short mECH) that removed the final helix, consistent with the boundaries used by Smith *et al*. Following the standard expression protocol, short mECH was expressed and purified. The yield was very low, the protein was poorly soluble and showed a tendency to precipitate out of solution. Therefore, it was decided not to continue working with this construct.

To confirm if mECH was folded, several biophysical assays were applied to probe the structure of the protein. Firstly, analytical gel filtration was conducted to understand more about its structure and oligomerisation in solution. The retention time did not suggest a defined oligomeric state as determined by poor fit to the calibrants (Figure 40 B and C). The retention time suggests the size of protein may be between a monomer and dimer. A circular dichroism (CD) spectrum was recorded for mECH and it showed a lack of defined peak shape that would be expected for a folded protein (Figure 40D). mECH is predicted to have the classic ββα crotonase superfamily fold, therefore negative peaks at 208 nm and 222 nm would be expected for the α-helices and a negative peak at 218 nm for anti-parallel β-strand.[138] However, an ill-defined curve is observed, suggesting that the does not contain defined secondary structure.

Finally, a ^{15}N labelled sample of mECH was prepared for high-resolution solution NMR analysis. Cells were cultured in LB medium to obtain cellular density prior to media exchange into a minimal media supplemented with $^{15}NH_4Cl$ as the nitrogen source. Ni^{2+}-affinity chromatography and SEC furnished a protein sample in NMR buffer (pH 7.5) that was concentrated to 400 µM and analysed by NMR (700 MHz). ^{1}H-^{15}N HSQC NMR revealed a lack of dispersity and broadness of signals that is the signature of an unfolded protein (Figure 40A). Coupled with the lack of functional activity, it appears it was not possible to reconstitute an active modular ECH domain as an individual enzyme.

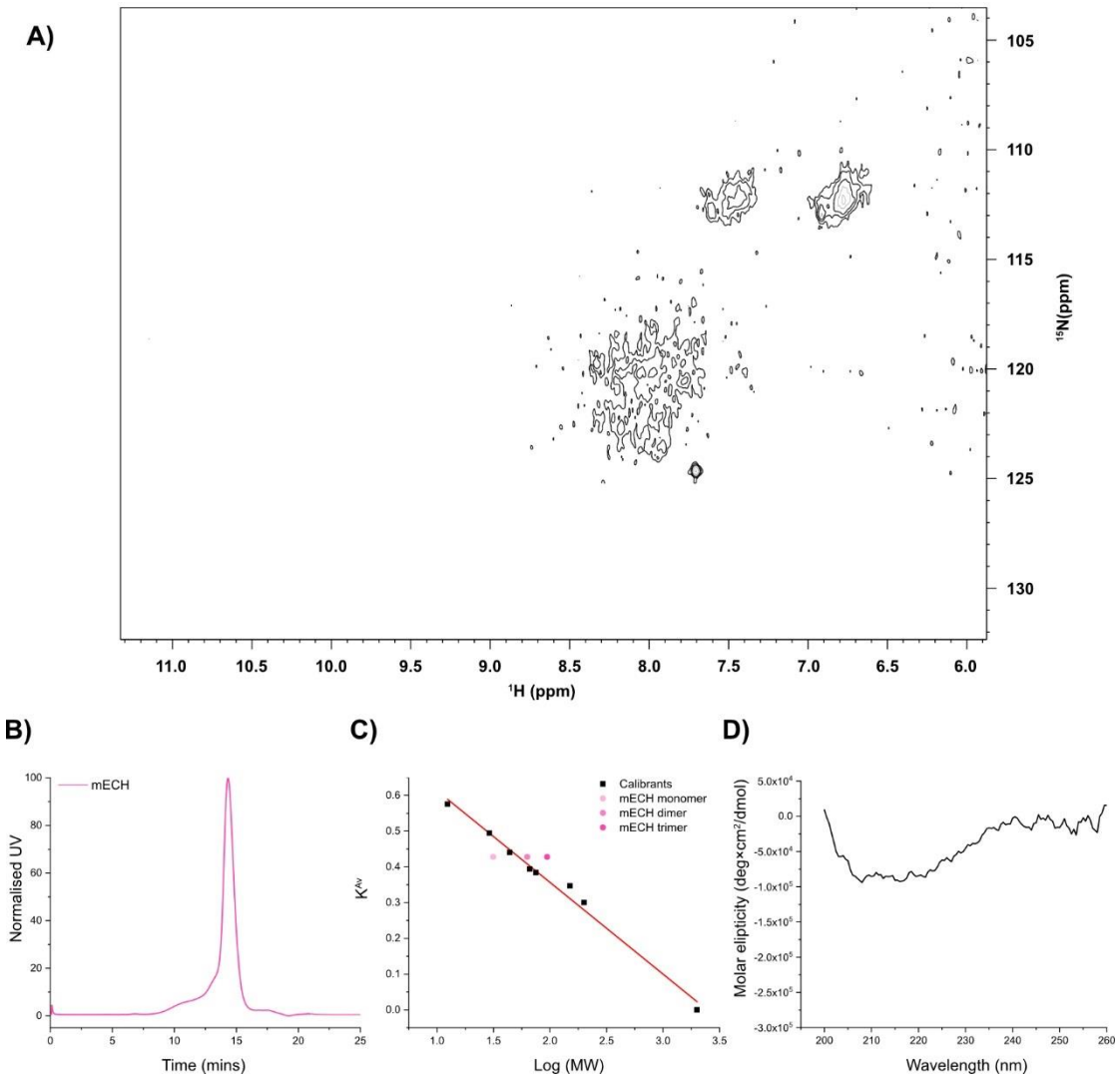


Figure 40 Structural characterisation of mECH. A) ^{1}H-^{15}N HSQC NMR spectra of [^{15}N]-labelled mECH. B) Analytical gel filtration of mECH (13.42 mins) run on a calibrated S200 column. C) Fitting of the retention time for mECH from the analytical GF indicates that the domain may be monomeric or dimeric, but a poor fit is observed. D) Circular dichroism spectrum for mECH does not indicate the expected α-helical dip at 208 nm and 222 nm or the β-strand dip at 218 nm.[138]

2.2.2 4M expression and structural analysis

Excising a single, functional domain from a large multi-modular protein is not straightforward.[139] As has been shown for mECH, a soluble protein could be isolated in high yield, however, only upon close structural analysis was the lack of folding observed. It was hoped that expression of mECH as a multi-domain protein with either an N-terminal or C-terminal partner might stabilise the correctly folded structure and allow for assay work to be undertaken.

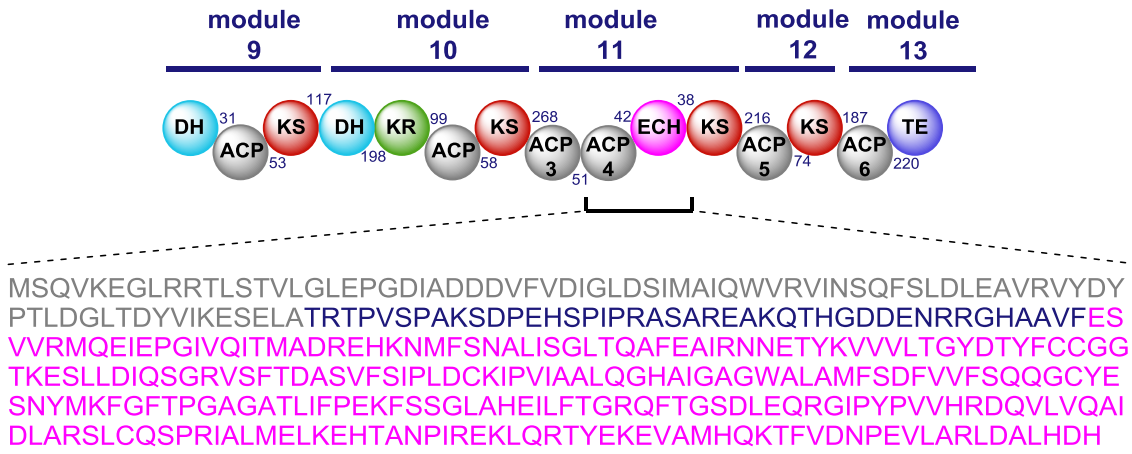


Figure 41 Order of domains and inter-domain linker lengths in Bat3 and the sequence of the di-domain 4M. ACP4 is shown in grey, the linker region in blue and the mECH in pink.

The order and length of linker of the domains in Bat3 is shown in Figure 41. ACP4 precedes the modular ECH domain and the two domains are separated by a 42-residue linker that is predicted to be unstructured. It was decided to express a di-domain comprised of ACP**4**-**m**ECH (**4M**) that preserved this linker. Primers were designed, the gene amplified and cloned into POPINF. Using the optimised procedure for the expression of ACP4, cells were grown, induced with 1 mM IPTG and incubated for 5 h at 16 °C. Following Ni^{2+}-affinity chromatography, 4M was isolated and confirmed as apo-4M by MS (Figure 42).

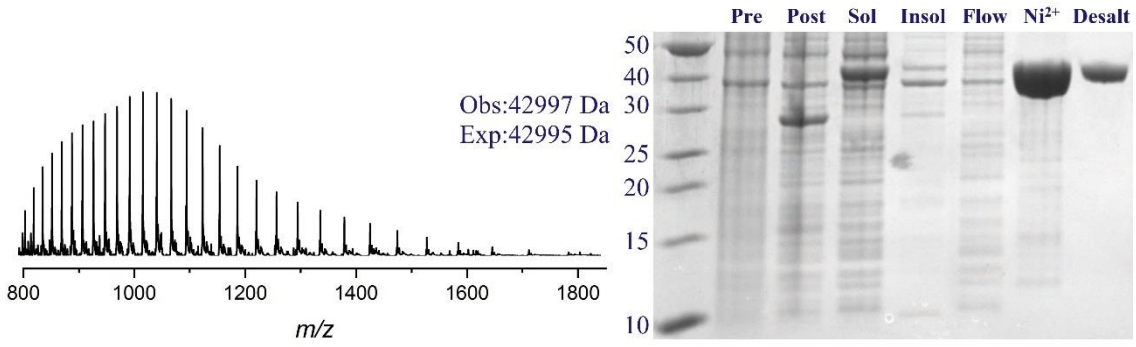


Figure 42 SDS-PAGE and MS analysis of the di-domain protein 4M (ACP**4**-**m**ECH). A single species is observed with good mass accuracy.

Prior to functional studies, structural characterisation of the protein was undertaken. Analytical gel filtration of 4M gave a species at 13.42 mins that correlated with the mass of a dimeric species (Figure 43 C and D). The crotonase superfamily predominantly form trimers or dimers of trimers, whereas structural studies on whole PKS modules show these multi-functional proteins form dimers around the KS domain.[33,35,100,140] The modular

ECH_2 CurF domain from the curacin pathway was shown to be trimeric in the solid state, but an equilibrium with lower oligomeric states existed in solution.[55] A CD spectrum of 4M contained dips at 208 nm and 222 nm, consistent with an α-helical protein (Figure 43E). It was not possible to determine the individual contribution to this signal from the α-helical ACP and the mECH domain which contains a mixed α-helical/β-strand structure. However, these results indicate that the 4M di-domain may be folded.

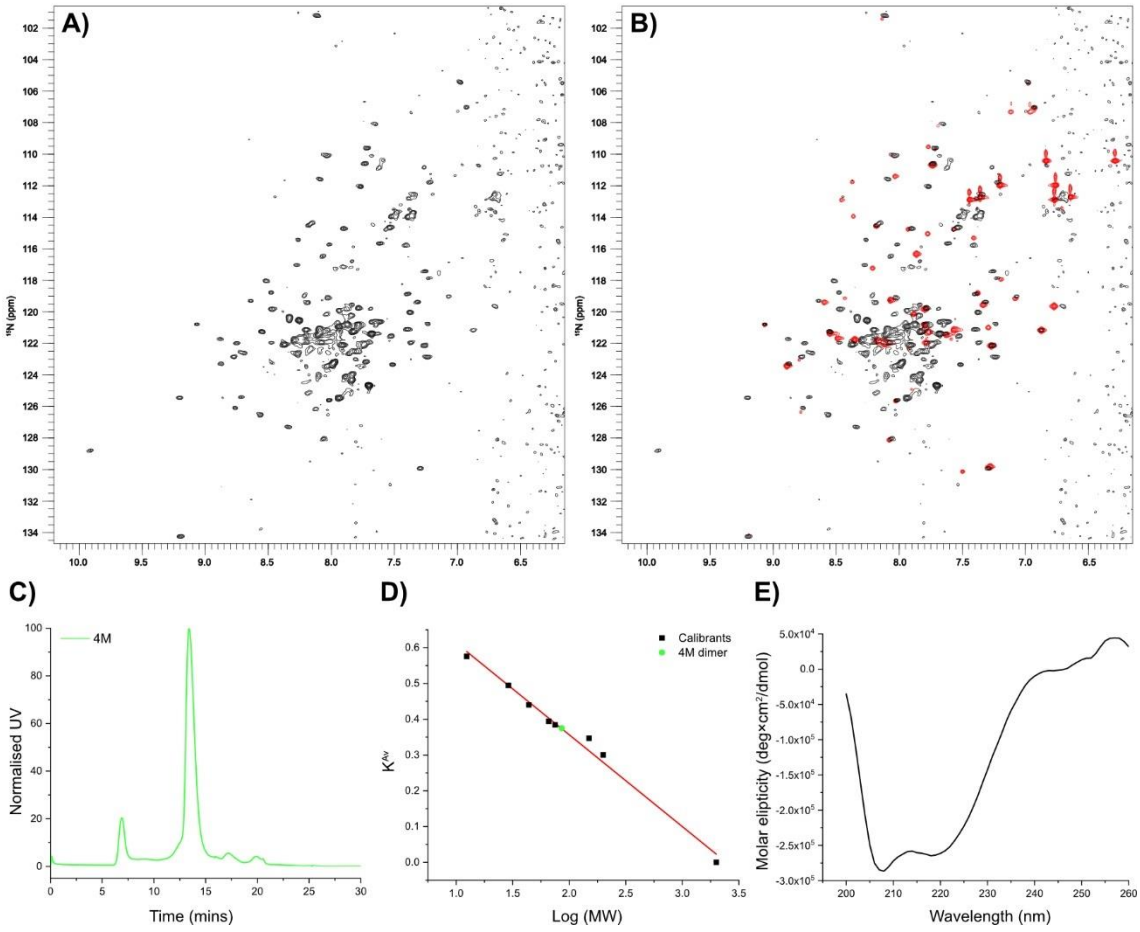


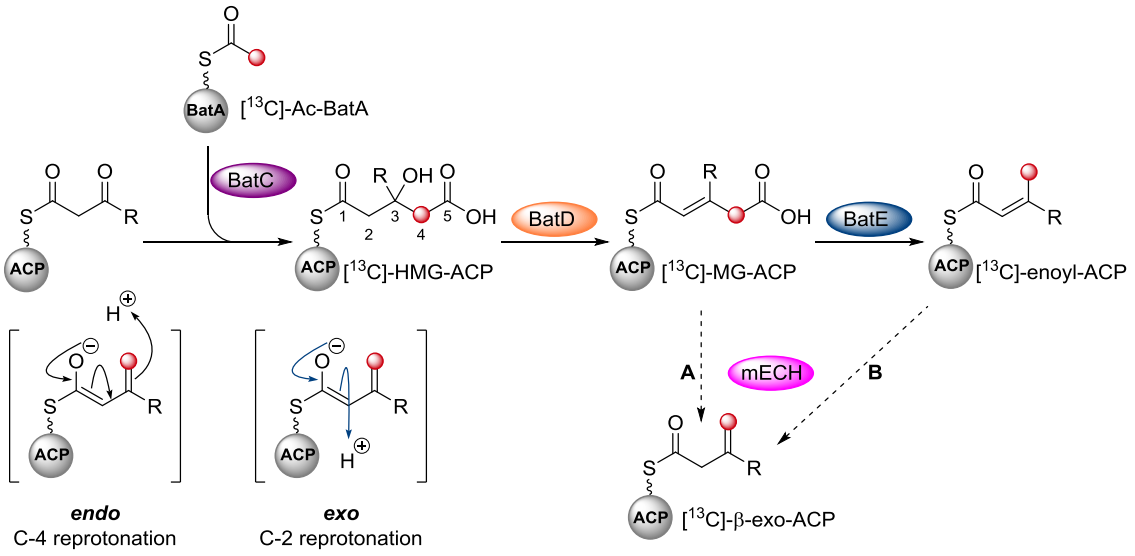
Figure 43 Structural characterisation of 4M. A) ^{1}H-^{15}N HSQC spectrum for 4M shows a greater dispersion of peaks compared to mECH. B) Overlay of ^{1}H-^{15}N HSQC spectra for 4M (black) and ACP4 (red) shows overlap of the ACP peaks as well as additional peaks arising from either the linker region or mECH domain. C) S200 analytical gel filtration of 4M. D) The retention time suggests that 4M is dimeric in solution. E) CD spectrum of 4M contains dips at 208 nm and 222 nm consistent with an α-helical protein.

To further probe the structure, a ^{15}N labelled sample was prepared in minimal media and analysed by 1D and 2D NMR. The ^{1}H NMR showed a series of characteristic peaks for ACP4 that had been previously observed. ^{1}H-^{15}N HSQC NMR showed a much greater dispersion and number of signals, although still less than expected, which might be related to the large mass of mECH and broad signals (Figure 43A). When the spectrum was

overlaid with that of ACP4 many of the signals for ACP4 could be mapped onto the di-domain spectrum, but numerous additional peaks could be resolved. These were surprisingly well dispersed and may arise from the linker region, which may be showing some structuring, as well as more mobile parts of the mECH (Figure 43B). The spectrum also lacked the concentration of broad, overlapping signals between 7.5-8.5 ppm characteristic of unfolded proteins. However, these may be swamped by the strong ACP4 and linker signals. Nonetheless, this gave a positive indication that 4M might be folded and was carried forward into functional studies.

2.2.3 Synthesis of [^{13}C]-labelled materials

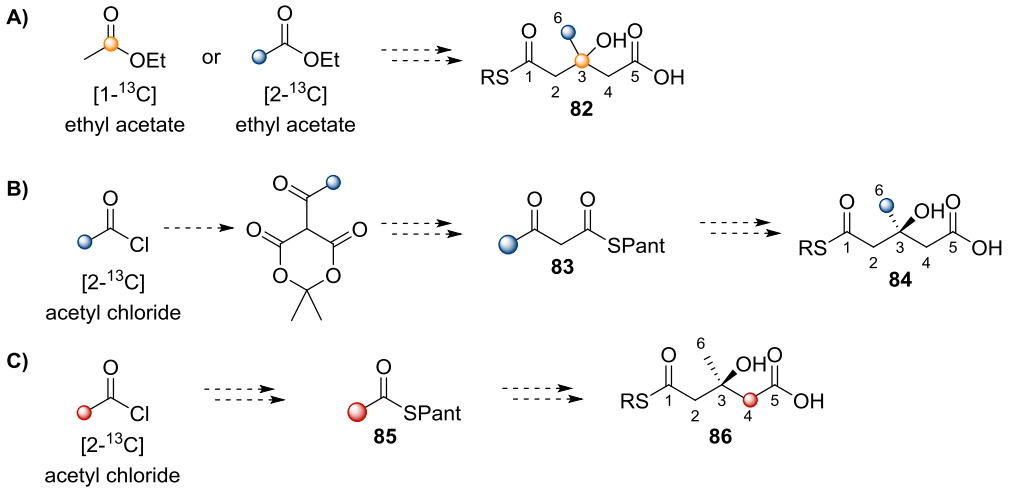
The success of *in vitro* NMR assays using HMG-pantetheine coupled with BatD/BatE to form an *endo*-β-methyl branch showed the potential for further development. An NMR assay that could analyse intermediates attached to an ACP or cleavage of intermediates from an ACP was required. Smith *et al.* demonstrated the conversion of thioesters from ACPs to give stable amide derivatives that were analysed by GC-MS and compared to synthetic standards.[45] Our aim was to create an NMR assay to analyse intermediates attached to the ACP as this would yield more information about the species formed during the β-branching pathway. Additionally, *in situ* analysis removes uncertainty associated with degradation or non-enzymatic side-reactions during hydrolysis and work-up.



Scheme 25 β-branching pathway showing the position of a [^{13}C] label (red dot) during the assay. The [^{13}C] label would occupy a range of chemical environments allowing for clear discrimination of β-branching intermediates.

It was envisioned that the incorporation of a [^{13}C] label, specifically introduced at one position in a β-branching substrate, would provide the requisite handle for *in situ* analysis

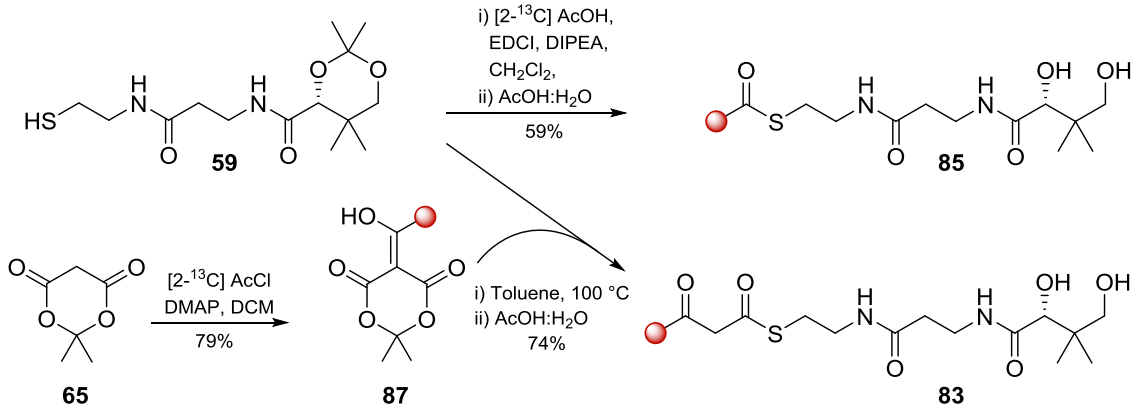
(Scheme 25). Monitoring reactions by direct observe ^{13}C NMR would allow for the observation of the enriched [^{13}C] label, whilst the 1% natural abundance ^{13}C of the ACPs, catalytic enzymes and other unlabelled components in the assay would not be observed.



Scheme 26 Potential positions of [^{13}C] incorporation into HMG-intermediates (**82**, **84** and **86**) where R may be pantetheine or ACP. A) [1-^{13}C] or [2-^{13}C] ethyl acetate would give rise to **82** containing a single [^{13}C] label at C-3 or C-6. B) [2-^{13}C] acetyl chloride could give **84** with single a [^{13}C] label at C-6 *via* [^{13}C]-Acac-pantetheine **83**. C) [2-^{13}C] acetyl chloride could be used to give **86** with a [^{13}C] label at C-4 *via* [^{13}C]-Ac-pantetheine **85**. However, **84** and **86** would require the BatC-catalysed condensation of ACP_A and ACP_D whereas **82** could be directly loaded onto ACP_A.

The position of the [^{13}C] label was key in designing the experiment, and three possibilities were envisioned. Firstly, utilising the previously undertaken synthetic route, [^{13}C]-HMG-intermediate **82** with a [^{13}C] label at C-3 or C-6 could be synthesised from 1- or 2-[^{13}C] ethyl acetate (Scheme 26A). The advantage of this intermediate is that it would negate the need for the BatC-catalysed condensation of two ACPs, thus simplifying the assay. However, the chemical environment of C-6 does not change significantly during β-branching and the previously used synthetic route to HMG-pantetheine **79** resulted in epimers at C-3, limiting the information that could be gained by this method.

The synthesis of [^{13}C]-Acac-pantetheine **83** or [^{13}C]-Ac-pantetheine **85** could be achieved using the previously developed synthetic routes from [2-^{13}C] acetyl chloride, a readily available and cost-effective source of [^{13}C]. Following BatC-catalysed condensation, [^{13}C]-Acac-pantetheine **83** would introduce a [^{13}C] label at C-6 of HMG-intermediate **84**, whereas [^{13}C]-Ac-pantetheine **85** would locate the [^{13}C] label at C-4 of HMG-intermediate **86** (Scheme 26B and C). Analysis of the mechanism of β-branching suggested a greater difference in chemical environment would be experienced by C-4 than C-6, making [^{13}C]-Ac-pantetheine **85** the more favourable biosynthetic mimic.



Scheme 27 Synthesis of [^{13}C]-Ac-pantetheine **85** and [^{13}C]-Acac-pantetheine **83**.

[^{13}C]-Ac-pantetheine **85** and [^{13}C]-Acac-pantetheine **83** were both prepared (Scheme 27). The synthesis of [^{13}C]-Acac-pantetheine **83** utilised [2-^{13}C] acetyl chloride analogous to the preparation of the unlabelled material. A higher yield of [^{13}C]-Ac-pantetheine **85** was obtained from the EDCI-mediated coupling of **59** with [2-^{13}C] acetic acid (performed by Angus Weir) rather than the condensation with [2-^{13}C] acetyl chloride.

The ^{1}H and ^{13}C NMR for both unlabelled (**60)** and [^{13}C] labelled (**85)** Ac-pantetheine are shown in Figure 44, and unlabelled (**62)** and [^{13}C] labelled (**83**) Acac-pantetheine in Figure 45. In the ^{1}H spectrum of [^{13}C]-Ac-pantetheine **85**, the coupling between the [^{13}C] enriched methyl carbon and the protons at 2.34 ppm can be observed with a $^{1}J_{H\text{-}C}$ coupling constant of 130 Hz (Figure 44B). The ^{13}C NMR showed an enriched signal at 30.5 ppm corresponding to the acetyl CH_3 (Figure 44C and D). The ^{1}H and ^{13}C spectra for [^{13}C]-Acac-pantetheine **83** are more complex due to the enol/keto tautomers formed. As such, the two methyl signals (1.94 ppm and 2.23 ppm) show $^{1}J_{C\text{-}H}$ coupling constants of 128 Hz, to give four signals in the spectrum (Figure 45B). In the ^{13}C spectrum, enhancement compared to the unlabelled material was seen for the methyl groups at 30.5 ppm and 20.8 ppm corresponding to the keto and enol forms respectively (Figure 45C and D).

Hence it was decided that [^{13}C]-Ac-pantetheine **85** was likely to be a better candidate for the proposed β-branching NMR assay. This was due to the single chemical shift and the more favourable C-4 position for the [^{13}C] label in an HMG-ACP intermediate following reaction of two ACPs. As a result, unlabelled Acac-pantetheine **62** was used as the β-ketothioester mimic attached to ACP_A, and [^{13}C]-Ac-pantetheine **85** was used as the acetyl donor tethered to ACP_D.

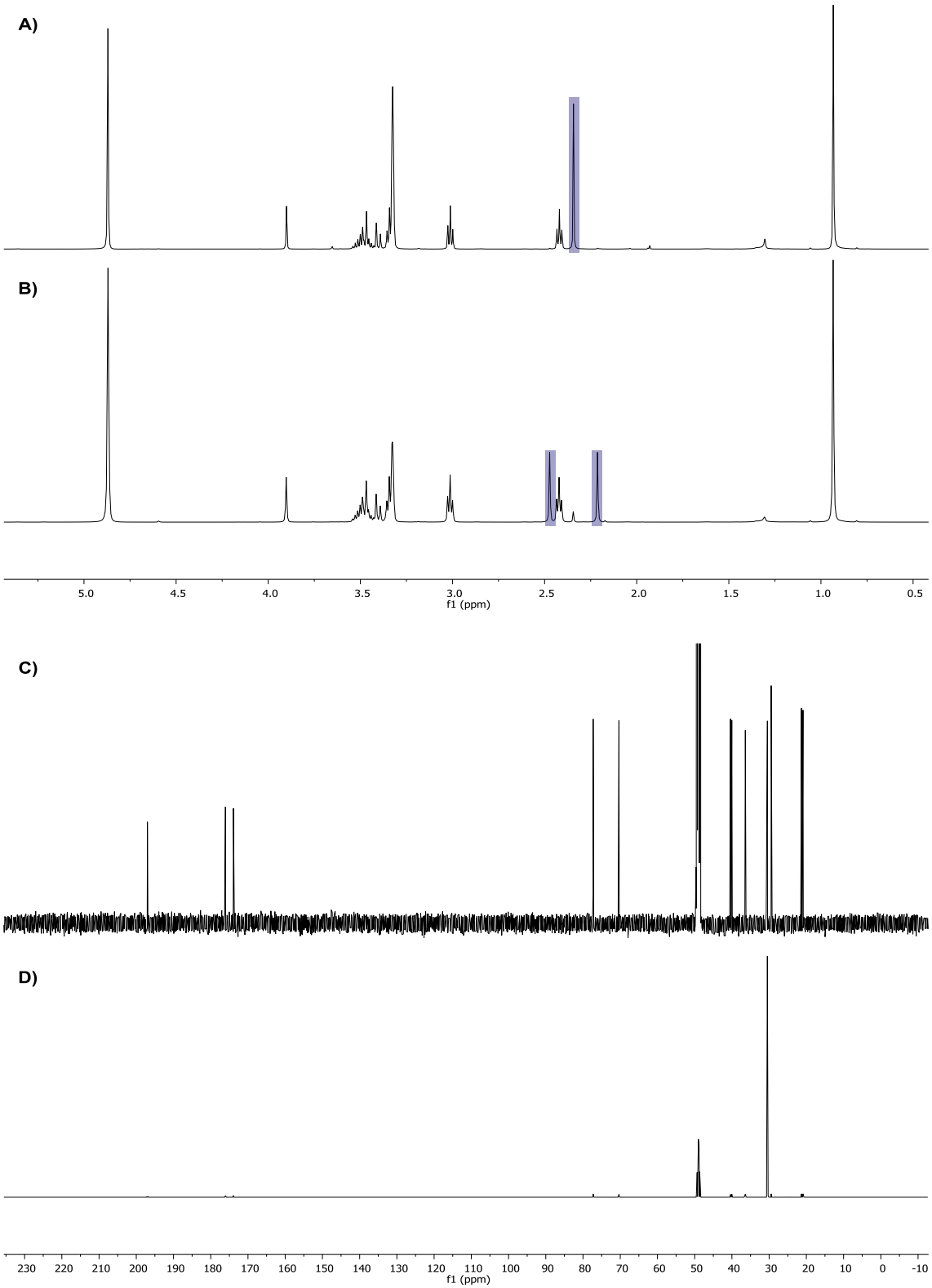


Figure 44 ^{1}H and ^{13}C NMR characterisation of unlabelled **60** and [^{13}C] labelled (**85**) Ac-pantetheine. A) ^{1}H spectrum of unlabelled Ac-pantetheine **60** with the acetyl CH_3 highlighted in blue. B) ^{1}H spectrum of [^{13}C]-Ac-pantetheine **85** shows coupling of the acetyl CH_3 signal (blue) at 2.34 ppm by $^{1}J_{H\text{-}C}$ coupling of 130 Hz. C) ^{13}C spectrum of unlabelled Ac-pantetheine **60**. D) ^{13}C spectrum of [^{13}C]-Ac-pantetheine **85** shows enhancement of the signal at 30.5 ppm due to the selective incorporation of the [^{13}C] label.

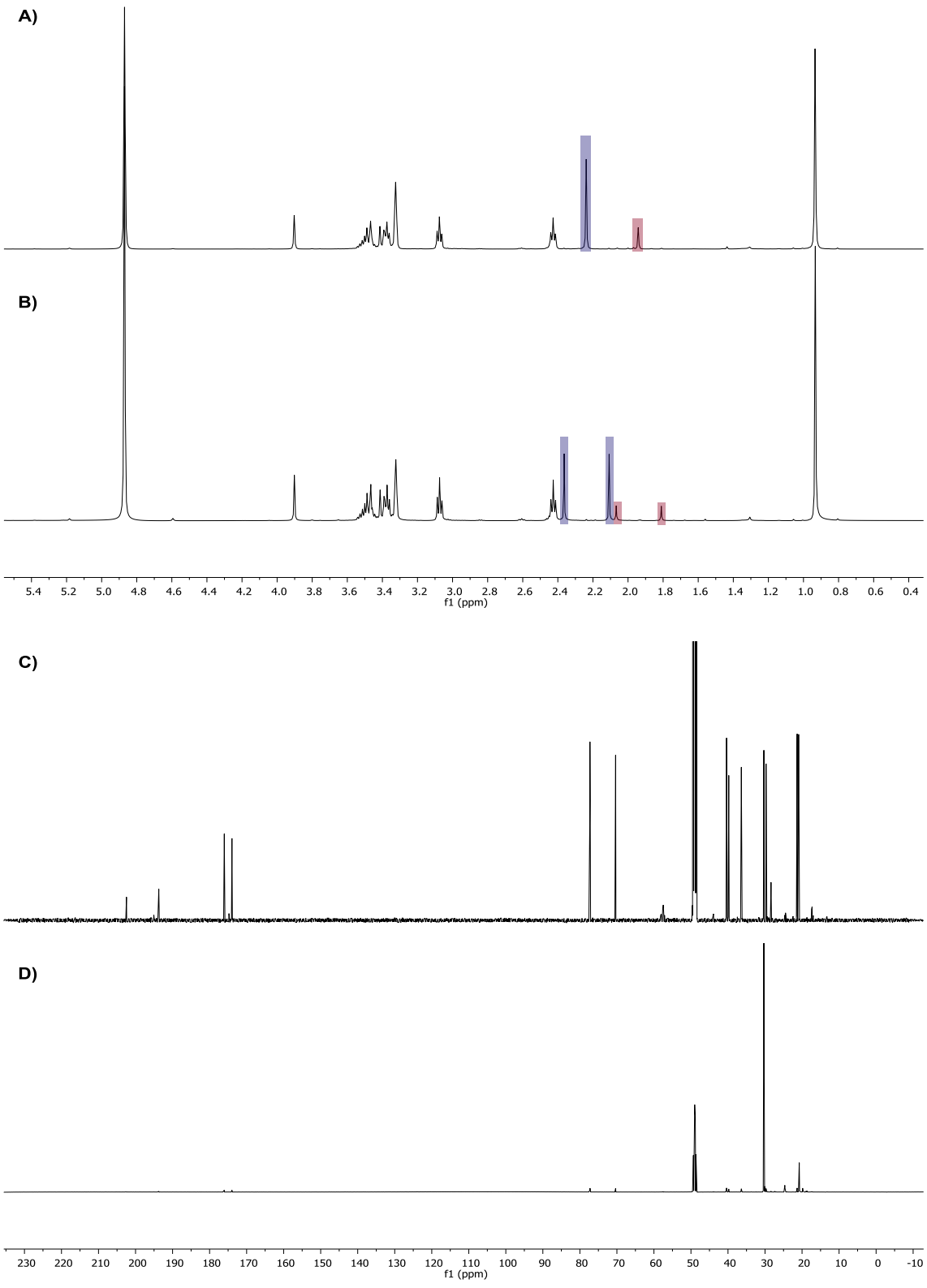


Figure 45 ^{1}H and ^{13}C NMR characterisation of unlabelled (**62**) and [^{13}C] labelled (**83**) Acac-pantetheine, with the major tautomer (keto) shown in blue and the minor (enol) tautomer shown in red. A) ^{1}H spectrum of unlabelled Acac-pantetheine **62**. B) ^{1}H spectrum of [^{13}C]-Acac-pantetheine **83** shows coupling of the keto-CH_3 at 2.23 ppm and the enol CH_3 at 1.94 ppm by a $^{1}J_{H\text{-}C}$ coupling of 128 Hz. C) ^{13}C spectrum of unlabelled Acac-pantetheine **62**. D) ^{13}C spectrum of [^{13}C]-Acac-pantetheine **83** shows enhancement of the signal at 30.5 ppm and 20.8 ppm due to the selective incorporation of the [^{13}C] label.

2.2.4 ^{13}C NMR assay for ACP4

With [^{13}C]-Ac-pantetheine **85** in hand, initial NMR experiments were carried out to determine whether the [^{13}C] label could be readily observed when attached to an ACP. Apo-BatA was loaded with [^{13}C]-Ac-pantetheine, as confirmed by MS analysis, desalted into NMR buffer to remove excess pantetheine and concentrated to 345 µM. A ^{1}H spectrum showed a dispersion of signals and upshifted methyl peaks were evident which arise from the interaction of methyl groups with aromatic side chains and is a good indication of folded protein (Figure 46A). A ^{13}C spectrum was recorded which yielded a single signal at 32.75 ppm (Figure 46B). The loading was monitored by MS and the Ppant ejection assay which gave a peak at 304.2 Da, 1 Da heavier than the unlabelled material.

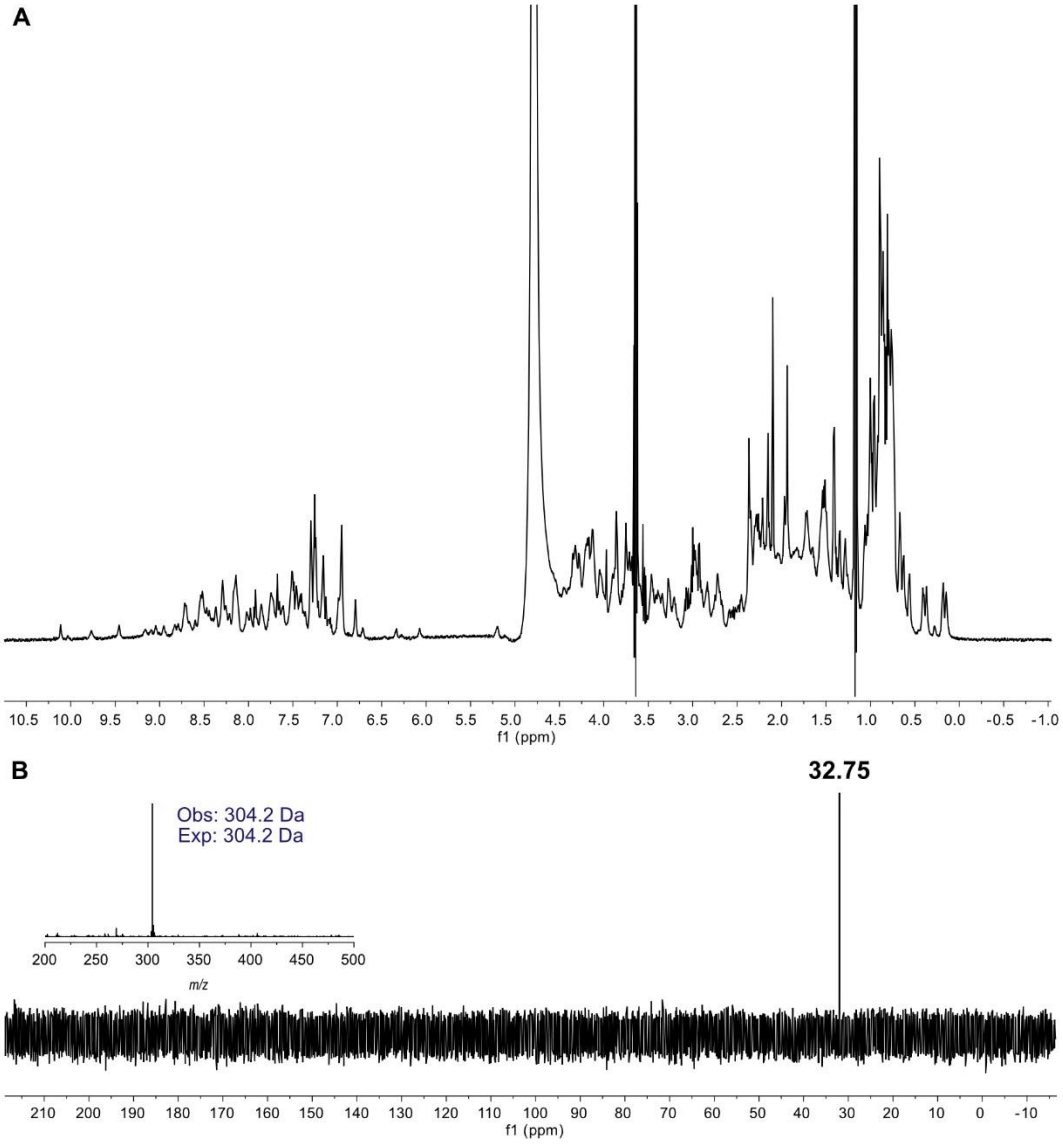


Figure 46 A) ^{1}H spectrum of [^{13}C]-Ac-BatA contains upfield shifted methyl peaks around 0.0-0.5 ppm which indicate protein folding. B) ^{13}C spectrum of [^{13}C]-Ac-BatA gave a single signal (32.75 ppm) and MS analysis by Ppant ejection (304,2 Da) confirmed the 1 Da increase in mass relative to unlabelled Ac-BatA (303.2 Da).

Before reconstituting the β-branching pathway, apo-ACP4 was converted to [^{13}C]-Ac-ACP4. Although a non-native substrate of the β-branching pathway, this control was run to ensure that the [^{13}C] label could be observed on this ACP by NMR. The loading was monitored by Ppant ejection and gave the expected peak at 304.2 Da. The sample was concentrated to 340 µM transferred to an NMR tube and a ^{1}H spectrum showed the protein was folded (Figure 47A). Once again, the ^{13}C spectrum showed a single signal at 32.75 ppm, which was identical to the chemical shift recorded for [^{13}C]-Ac-BatA (Figure 47B).

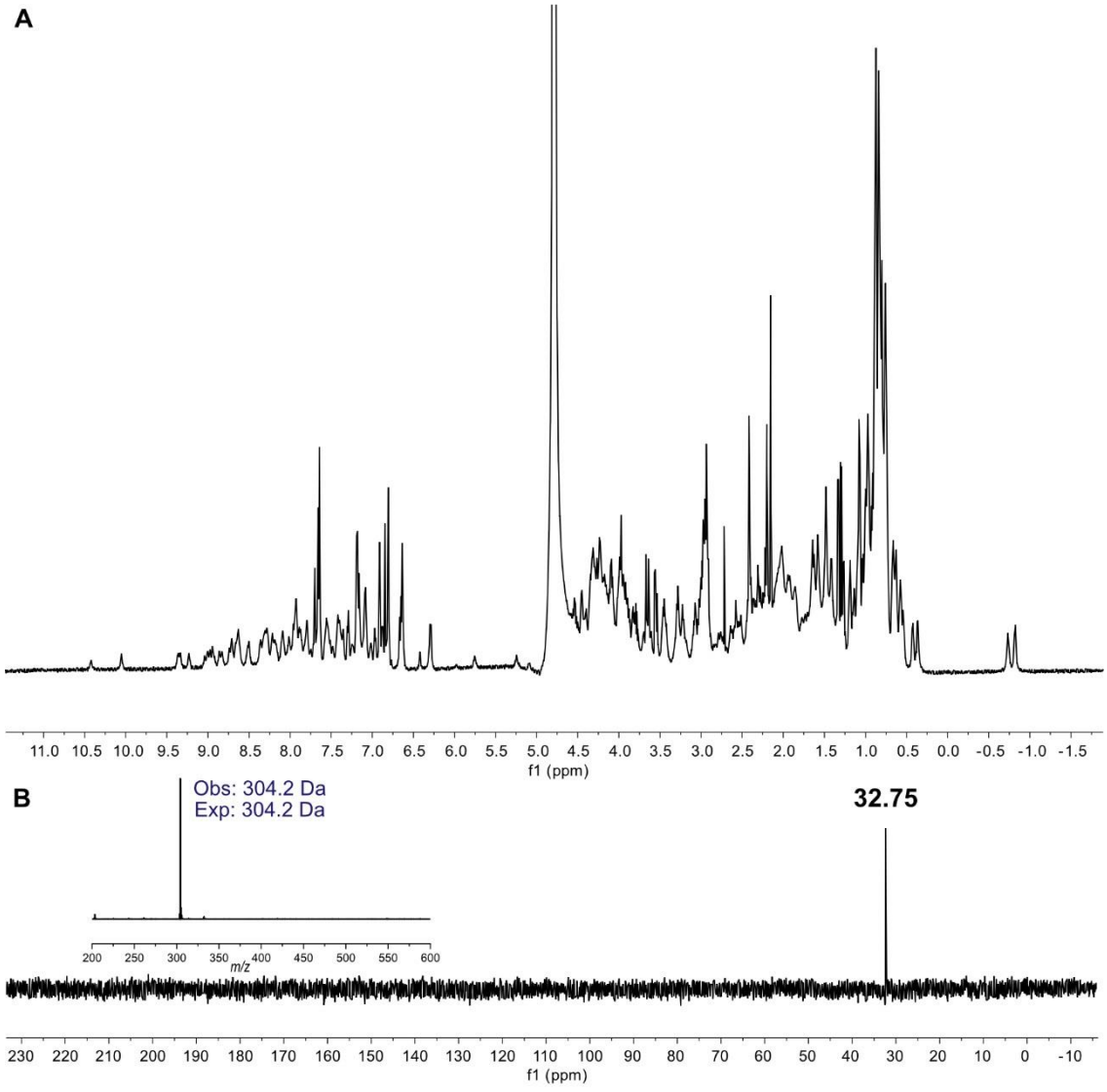


Figure 47 A) ^{1}H spectra of [^{13}C]-Ac-ACP4. B) ^{13}C spectrum of [^{13}C]-Ac-ACP4 gives a single signal at 32.75 ppm that, along with MS analysis by Ppant ejection (304.2 Da), confirmed the loading of [^{13}C]-labelled acetate on to ACP4.

The loading of BatA and ACP4 with [^{13}C]-Ac-pantetheine and NMR experiments provided information on the ease of loading, amount of [^{13}C] label required and the

sensitivity of the NMR experiments. The reaction of two ACPs results in sample dilution, therefore it was deemed necessary to get a relatively high stock concentration of [^{13}C] label for the NMR experiments. This would allow for shorter NMR experiments which would increase the chance of observing transient intermediates. During β-branching assays the sample was constantly monitored and NMR spectra recorded at appropriate times. Length and frequency of NMR experiments varied according to each individual experiment and the amounts of [^{13}C]-labelled products formed.

To test the methodology during the challenging reconstitution of the β-branching pathway, the assay was repeated on ACP4 as previously shown by MS and pantetheine-bound NMR experiments to give an *endo*-β-methyl branch. Apo-ACP4 was loaded with Acac-pantetheine **62**, desalted and concentrated. [^{13}C]-Ac-BatA and BatC were added and the reaction mixture transferred to an NMR tube. The ^{1}H NMR showed the characteristic upshifted methyl peaks for both ACPs (Figure 48).

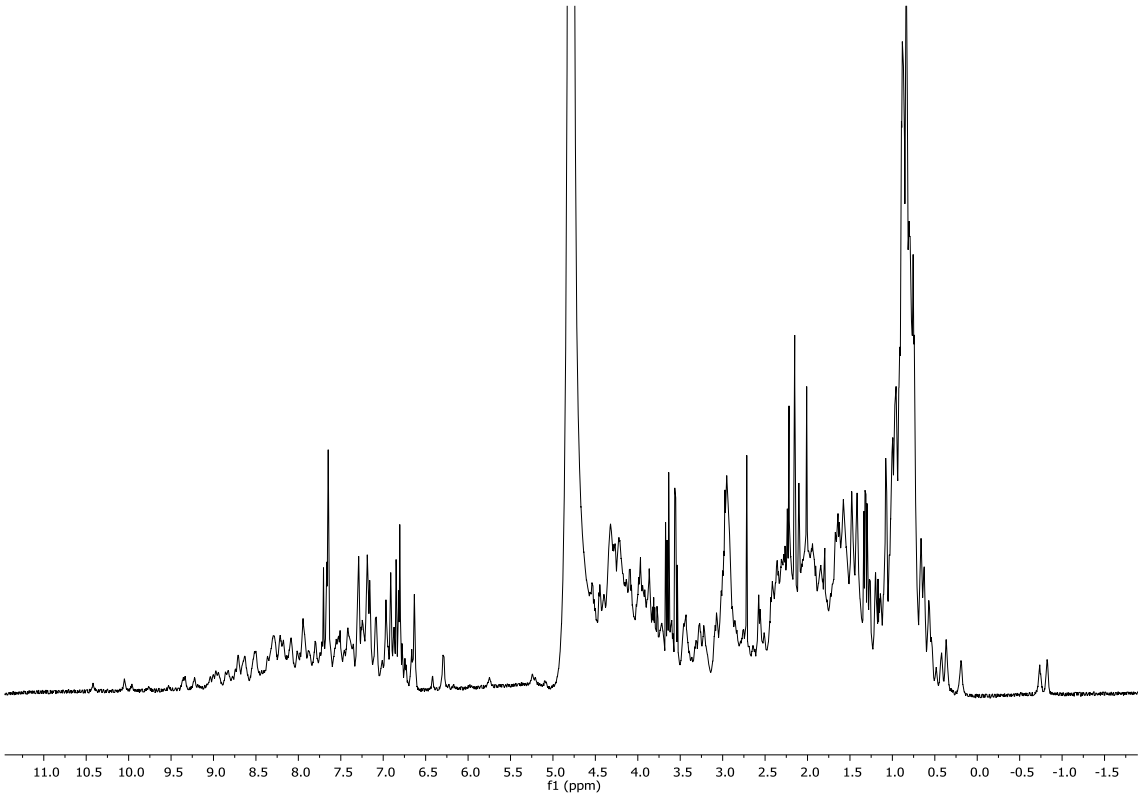


Figure 48 ^{1}H spectrum of Acac-ACP + [^{13}C]-Ac-BatA + BatC shows upfield shifted methyl peaks corresponding to both ACPs.

The ^{13}C spectra showed two previously uncharacterised peaks in addition to [^{13}C]-Ac-BatA at 32.75 ppm (Figure 49). The new peak at 50.15 ppm had a chemical shift consistent with the CH_2 of an ACP-bound HMG-intermediate and MS analysis showed

the formation of the HMG-intermediate plus 1 Da for the [^{13}C] label (406.2 Da), by Ppant ejection assay. A DEPT-edited ^{13}C spectrum was recorded and the peak at 50.15 ppm was inverted consistent with a CH_2 group, and adding further evidence to the successful formation of [^{13}C]-HMG-ACP4. The peak at 26.00 ppm remained positively phased in the DEPT spectra and was assigned as free [^{13}C]-acetate. [^{13}C]-Ac is transferred from BatA to BatC prior to condensation to form [^{13}C]-HMG-ACP. However, non-productive hydrolysis of [^{13}C]-Ac-BatC results in the formation of free [^{13}C]-acetate. This undesired reaction has been reported for the HMGCS of primary metabolism and CurD (HMGS) from the curacin β-branching cassette.[41,141] The sample was monitored over 16 h by recording back-to-back ^{13}C spectra and the complete hydrolysis of [^{13}C]-Ac-BatA to free [^{13}C]-acetate could be observed over time. The [^{13}C]-HMG-ACP4 peak at 50.15 ppm remained constant consistent with no further reaction progress or hydrolysis.

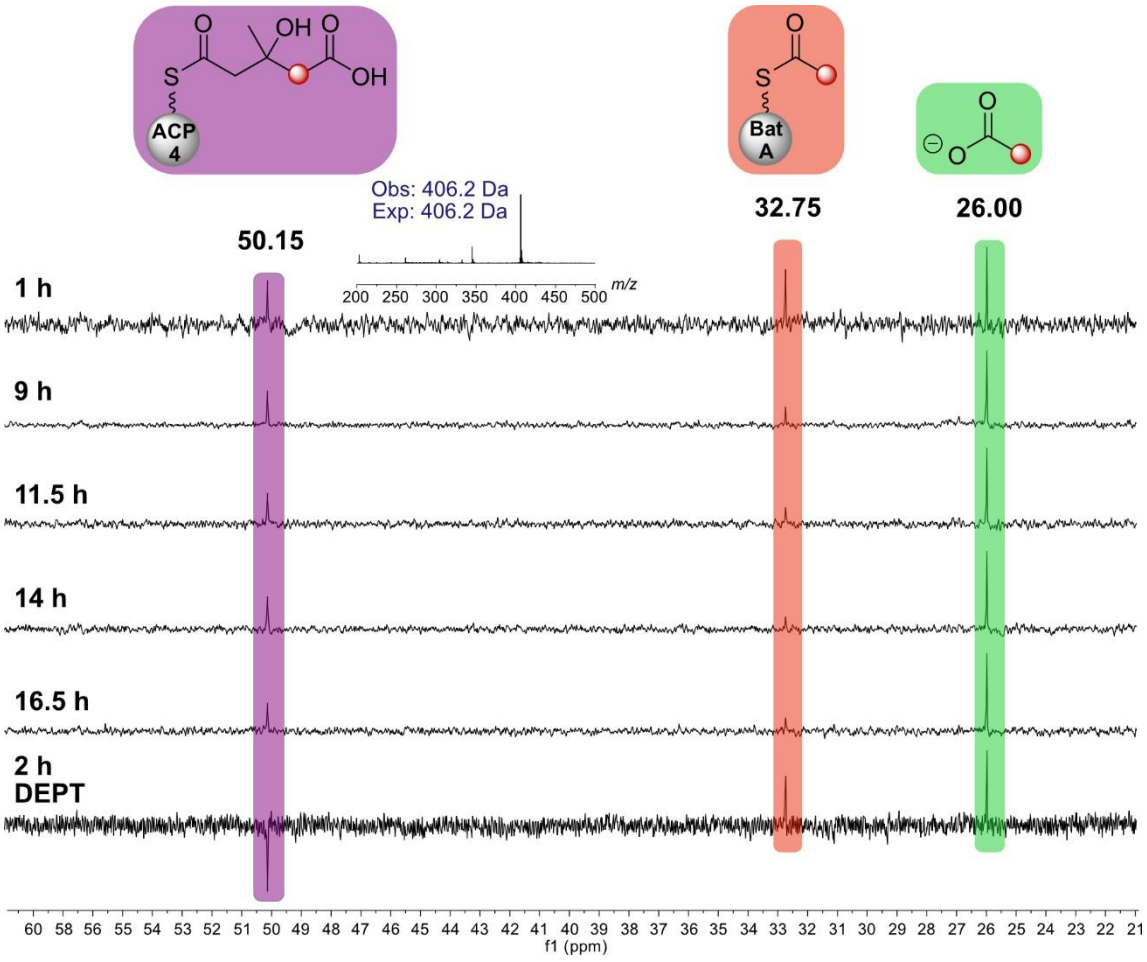


Figure 49 Stack of ^{13}C spectra of Acac-ACP4 + [^{13}C]-Ac-BatA + BatC with an initial spectrum taken at 1 h up to 16.5 h after initiation of the assay. A DEPT-edited spectrum resulted in the phase inversion of the peak at 50.15 ppm, corresponding to the CH_2 group of [^{13}C]-HMG-ACP4. MS analysis of the reaction after 1 h gave a peak at 406.2 Da by Ppant ejection assay, corresponding to [^{13}C]-HMG-ACP4.

To investigate the dehydration of [^{13}C]-HMG-ACP4, the intermediate was freshly generated by BatC, at which point BatD was added and a ^{1}H spectrum was recorded to confirm ACP folding (Figure 50).

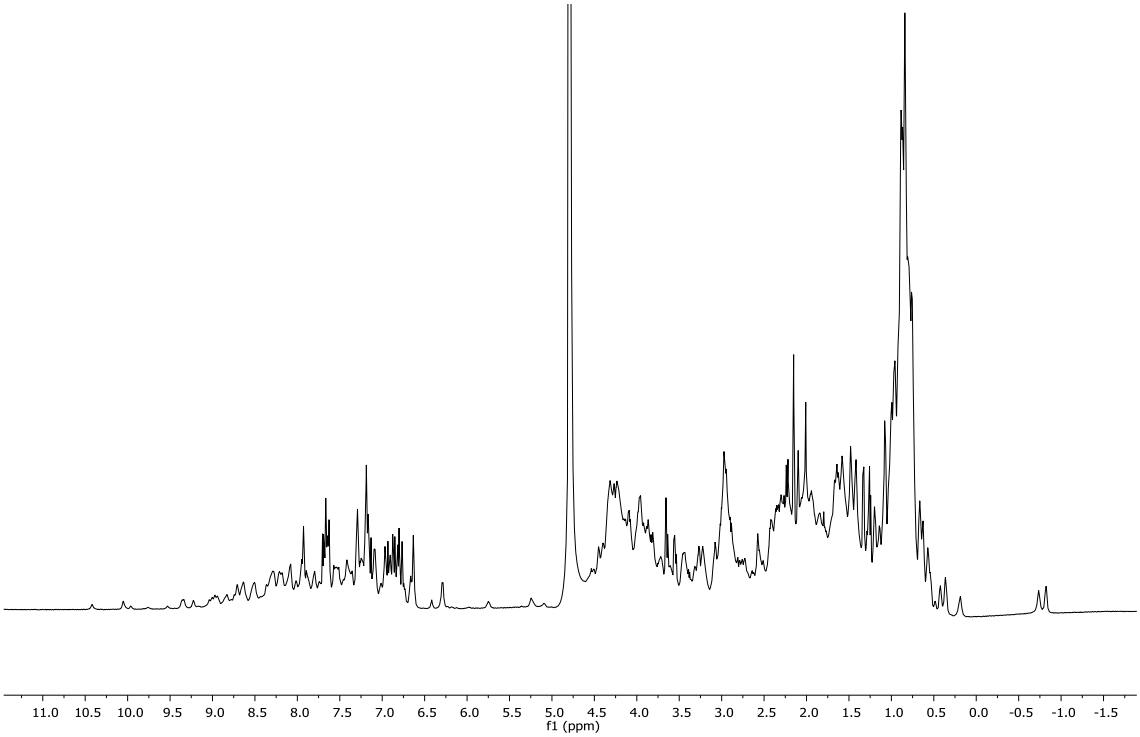


Figure 50 A ^{1}H NMR spectrum of Acac-ACP + [^{13}C]-Ac-BatA + BatC + BatD.

A DEPT spectrum was then recorded and showed the same three peaks corresponding to [^{13}C]-Ac-BatA, [2-^{13}C]-acetate and [^{13}C]-HMG-ACP4 (Figure 51). The reaction was monitored over 11 h, however, no new signal corresponding to the dehydrated intermediate [^{13}C]-MG-ACP4 was observed. The reversible nature of the reaction and the low yield of MG-pantetheine **80** from our previous experiments suggested that this reaction would be difficult to monitor. A 2D HSQC spectrum was recorded and showed the ^{1}H-^{13}C correlations for free [^{13}C]-acetate (26.00 ppm/1.91 ppm), [^{13}C]-HMG-ACP4 (51.15 ppm/2.42 ppm) and two newly identified peaks at 42.10 ppm and 52.1 ppm (Figure 52). The peak at 42.10 ppm was not observable in the DEPT spectrum, but due to the greater ^{1}H sensitivity of the NMR instrument used, it was possible to detect this species.

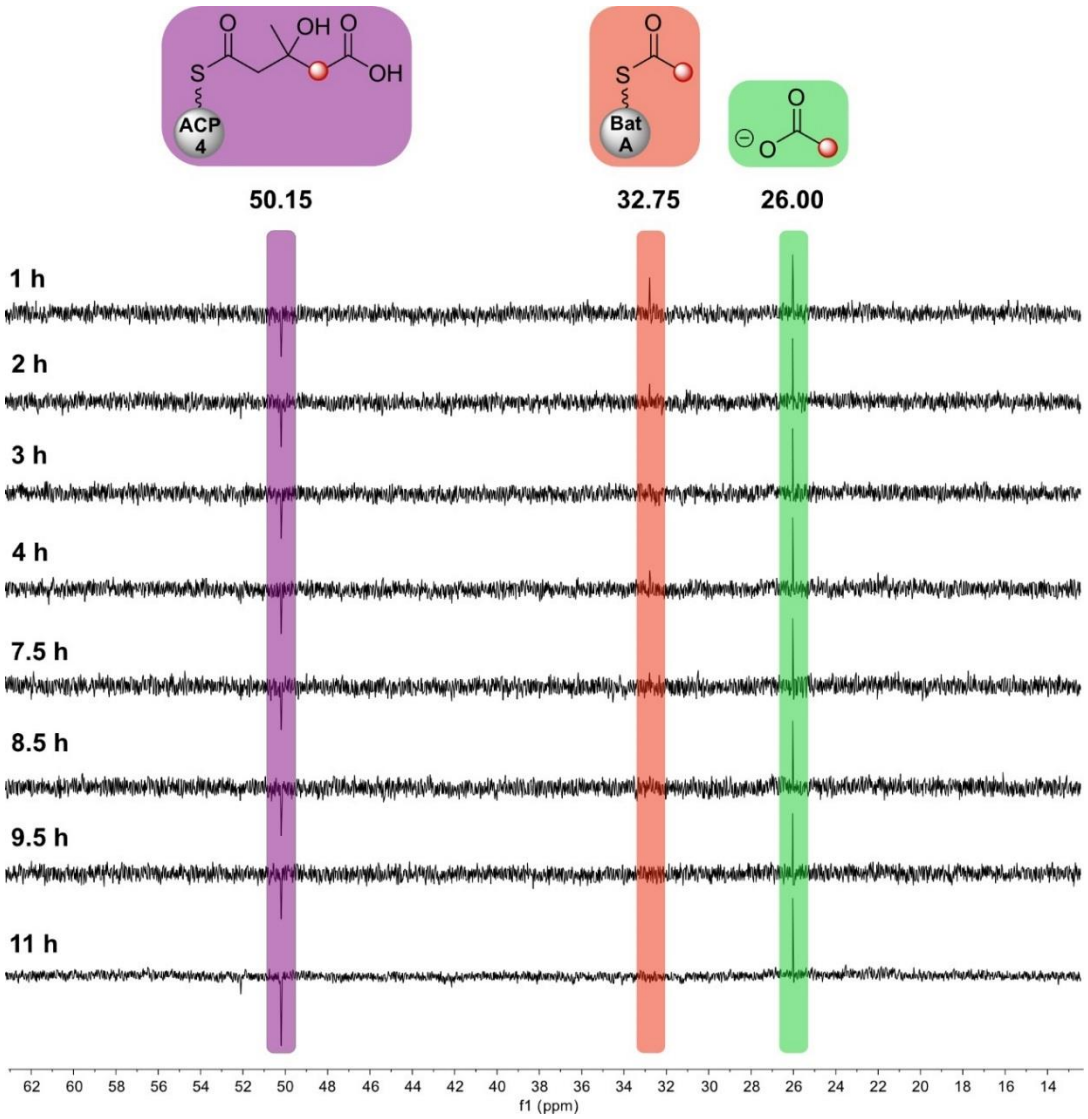


Figure 51 Stack of DEPT spectra showing the stability of [^{13}C]-HMG-ACP4 when incubated with BatD. No dehydrated product is observed, however, a new minor peak at 52.1 ppm appeared after 11h that could not be assigned.

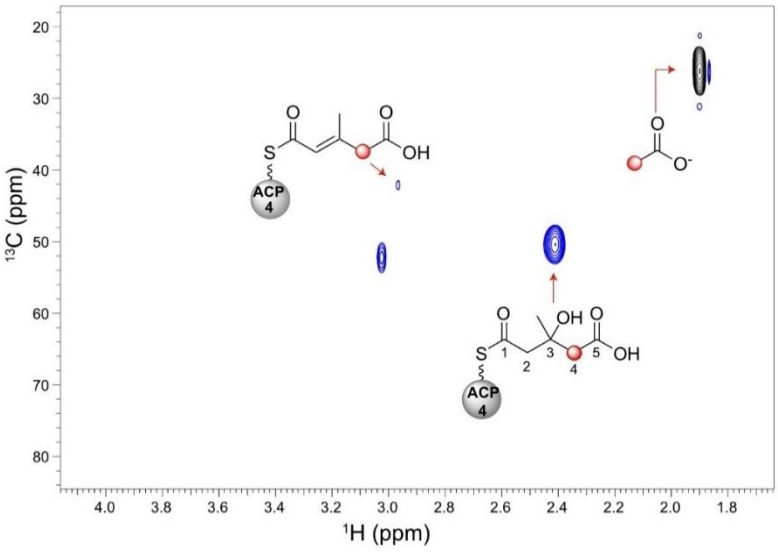


Figure 52 HSQC spectrum gives the ^{1}H-^{13}C correlations for [2-^{13}C]-acetate (26.00 ppm/1.91 ppm) and [^{13}C]-HMG-ACP4 (50.15 ppm/2.42 ppm). The peak at 42.10 ppm with a ^{1}H correlation to 2.97 ppm corresponds to [^{13}C]-MG-ACP4 The peak at 52.1 ppm could not be identified. The peaks shown in black are positively phased (CH or CH_3) and the blue peaks are negatively phased (CH_2).

Although hydrolysis was not observed in the previous BatC-condensation experiment, it was thought that one of the two unidentified species may correlate to HMG hydrolysis. The unlabelled diacid **77** had been previously synthesised and characterised in $(CD_3)_2CO$, so the compound was dissolved in the assay NMR buffer and ^{1}H, ^{13}C and ^{1}H-^{13}C HSQC spectra were recorded. The ^{13}C chemical shift of the CH_2 was 51.2 ppm and showed a correlation to a ^{1}H peak at 2.36 ppm (Figure 53A). Neither chemical shifts match the observed peak, ruling out hydrolysis of [^{13}C]-HMG-ACP4 and so the identity of the species at 52.1 ppm remained unknown.

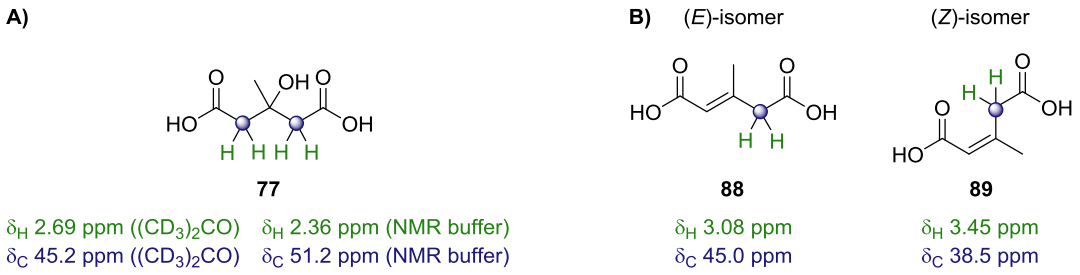


Figure 53 A) ^{1}H and ^{13}C chemical shifts of the hydrolysed di-acid in d_6-acetone and NMR buffer. B) Literature assignment in D_2O of the *E*- and *Z*- isomers of the dehydrated di-acid.[142]

The peak at 42.10 ppm, with a ^{1}H correlation at 2.97 ppm, corresponded to dehydration intermediate [^{13}C]-MG-ACP4. From the previous pantetheine experiments, the ^{1}H chemical shift of the CH_2 group of MG-pantetheine **80** was 3.07 ppm. As the intermediate is now ACP-bound, the slight change in chemical shift is not unreasonable. To further validate this result, a comparison of chemical shifts was made with the previously characterised di-acids **88** and **89** (Figure 53B).[142] The ^{13}C chemical shift for the CH_2 of the dehydrated di-acid in D_2O were 45.0 ppm (^{1}H correlation at 3.08 ppm) for the major *E*-isomer **88** and 38.5 ppm (^{1}H correlation at 3.45 ppm) for the minor *Z*-isomer **89**. Considering the difference in NMR solvent and the influence of the ACP attached *via* a thioester linkage, these data support the peak at 42.10 ppm corresponding to [^{13}C]-MG-ACP4.

Finally, [^{13}C]-HMG-ACP4 was freshly prepared from Acac-ACP4, [^{13}C]-Ac-BatA and BatC, at which point BatD and BatE were added and a DEPT spectra was recorded (Figure 54C). A new, positively phased peak at 23.51 ppm was observed in addition to the previously identified species. Ppant ejection yielded two ions corresponding to an unsaturated-β-methyl-ACP4 (344.2 Da) and Acac-ACP4 (345.2 Da).

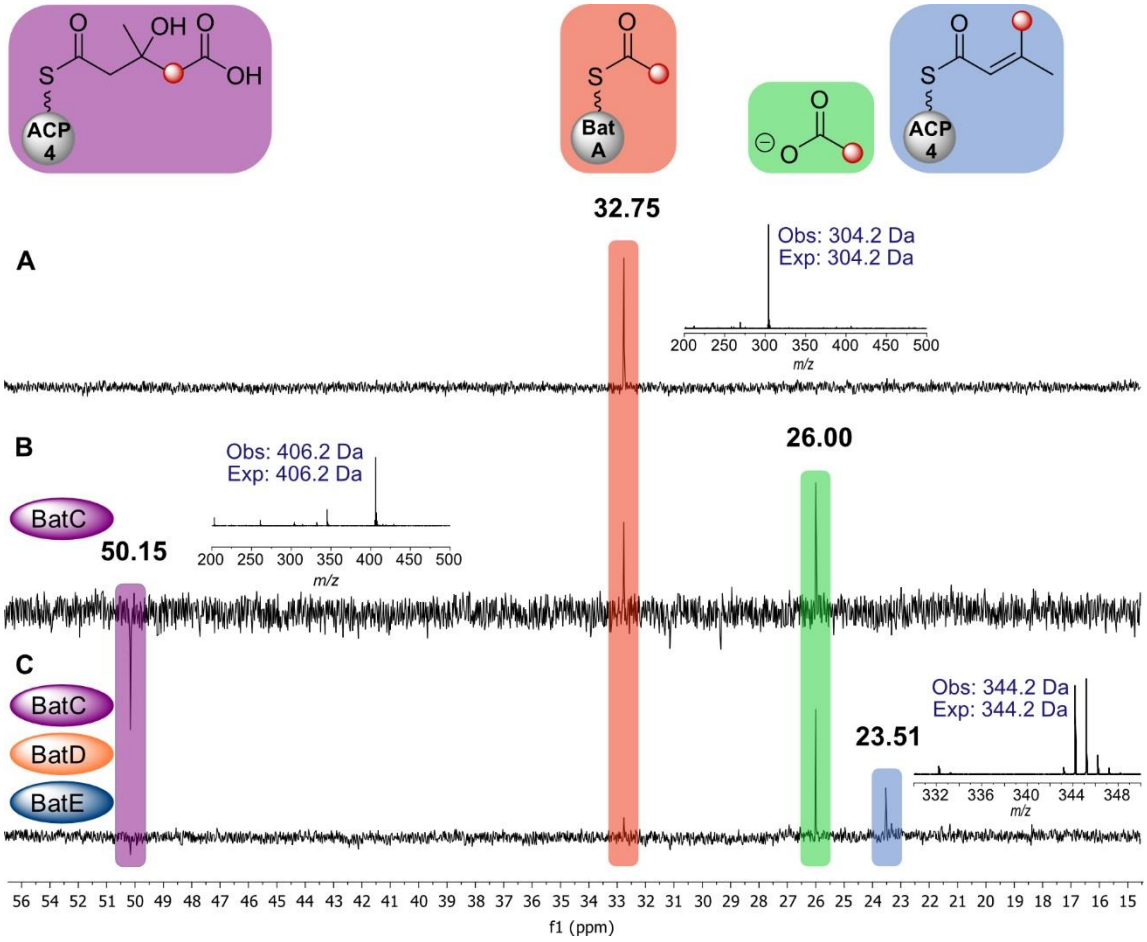


Figure 54 Stack of ^{13}C DEPT-edited spectra and MS analysis by Ppant ejection assay of the β-branching pathway for ACP4 to give an *endo*-β-methyl branch. A) [^{13}C]-Ac-BatA, B) Acac-ACP4 + [^{13}C]-Ac-BatA + BatC, C) Acac-ACP4 + [^{13}C]-Ac-BatA + BatC + BatD + BatE. The Ppant ejection gives two peaks corresponding to *endo*-β-methyl-ACP4 (344.2 Da) and Acac-ACP4 (345.2 Da), both arising from the ejection due to the 1 Da mass difference.

A ^{1}H-^{13}C HSQC spectrum was acquired and the [^{13}C] labelled CH_3 at 23.51 ppm showed a ^{1}H correlation at 2.03 ppm (Figure 55). The chemical shifts matched that of the previously analysed standards: enoyl-pantetheine **64** and enoyl-$^{13}C/^{15}N$-ACP4. It was possible to assign the [^{13}C] labelled methyl group *trans* to the alkene-proton with the unlabelled methyl group *cis*, as expected for the linearised polyketide chain, by comparison with the previously determined chemical shifts.

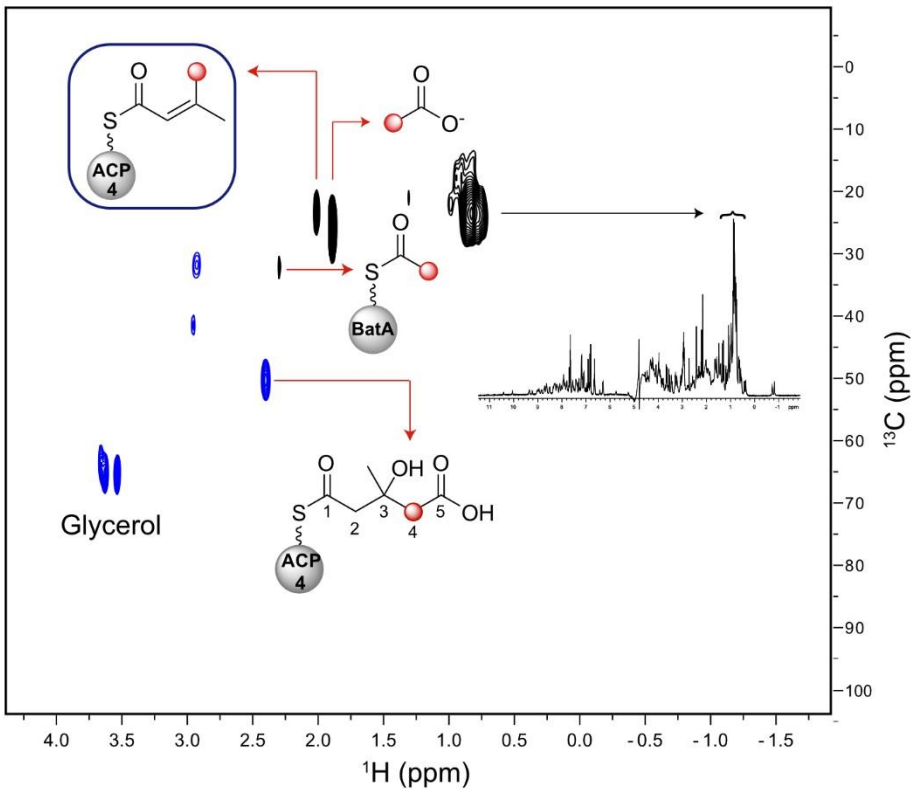


Figure 55 ^{1}H-^{13}C HSQC spectrum of the reaction Acac-ACP4 + [^{13}C]-Ac-BatA + BatC + BatD + BatE. The [^{13}C]-labelled methyl at 23.51 ppm shows a correlation to a proton at 2.03 ppm consistent with the formation of the *endo*-β-methyl product. [^{13}C]-HMG-ACP (50.15 ppm/2.42 ppm) and [^{13}C]-Ac-BatA (32.75 ppm/2.31) as well as [^{13}C]-acetate (26.00 ppm/1.91 ppm) arising from non-productive hydrolysis of [^{13}C]-Ac-BatC are also present in the spectrum. Peaks arising from the ACP backbone are visible between 0.5-1.0 ppm (insert).

With the formation of an *endo*-β-methyl on ACP4 confirmed, the pathway was reconstituted on an alternative β-branching ACP to demonstrate the generality of the methodology. ACP5 was loaded with Acac-pantetheine **62**, reacted with [^{13}C]-Ac-BatA in the presence of BatC and incubated with BatD and BatE. The ^{1}H spectrum for ACP5 once again showed a folded protein and the DEPT-spectra showed clean conversion to an *endo*-β-methyl in addition to free [2-^{13}C]-acetate (Figure 56). The chemical shifts of the product (23.49 ppm) were very similar to that of ACP4 (23.51 ppm), showing that different ACPs do not have a major impact on the chemical shift of the thioester-bound substrate. A ^{1}H-^{13}C HSQC spectrum revealed a greater difference in chemical shifts in the ^{1}H axis between ACP5 (1.95 ppm) and ACP4 (2.03 ppm) than observed in the ^{13}C axis (Figure 57). However, MS supported the formation of an unsaturated β-branch with Ppant ejection ion at 344.2 Da.

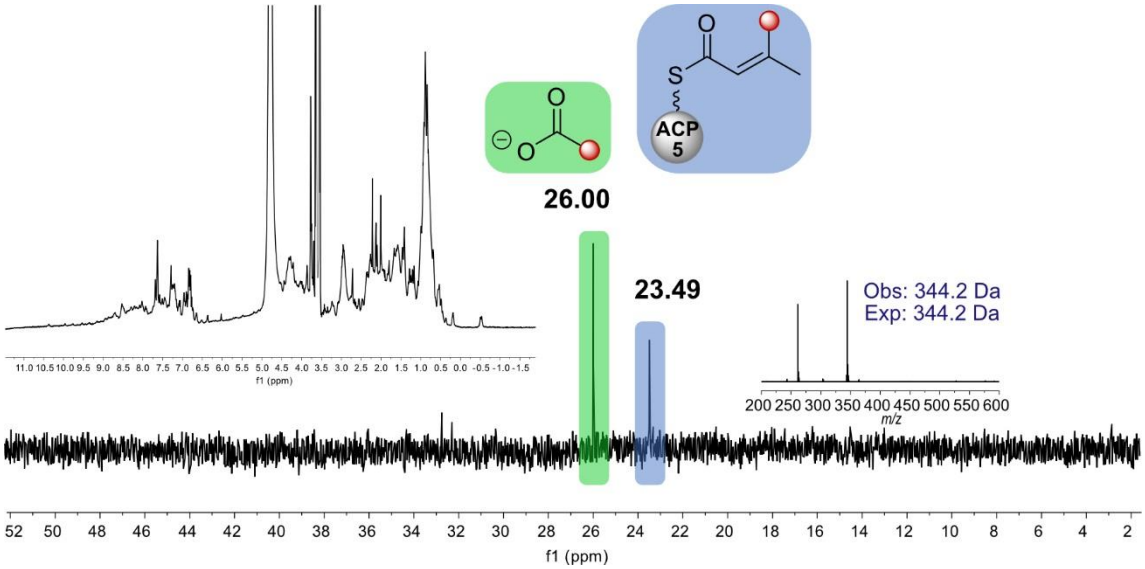


Figure 56 Characterisation of an *endo*-β-methyl branch attached to ACP5 from the reaction Acac-ACP5 + [^{13}C]-Ac-BatA + BatC + BatD + BatE. ^{13}C DEPT-edited spectra and MS analysis by Ppant ejection assay confirming the identity of the *endo*-β-methyl at 23.49 ppm and a mass of 344.2 Da by Ppant ejection. Insert: ^{1}H spectrum showing upfield shifted methyl peaks indictive of folding.

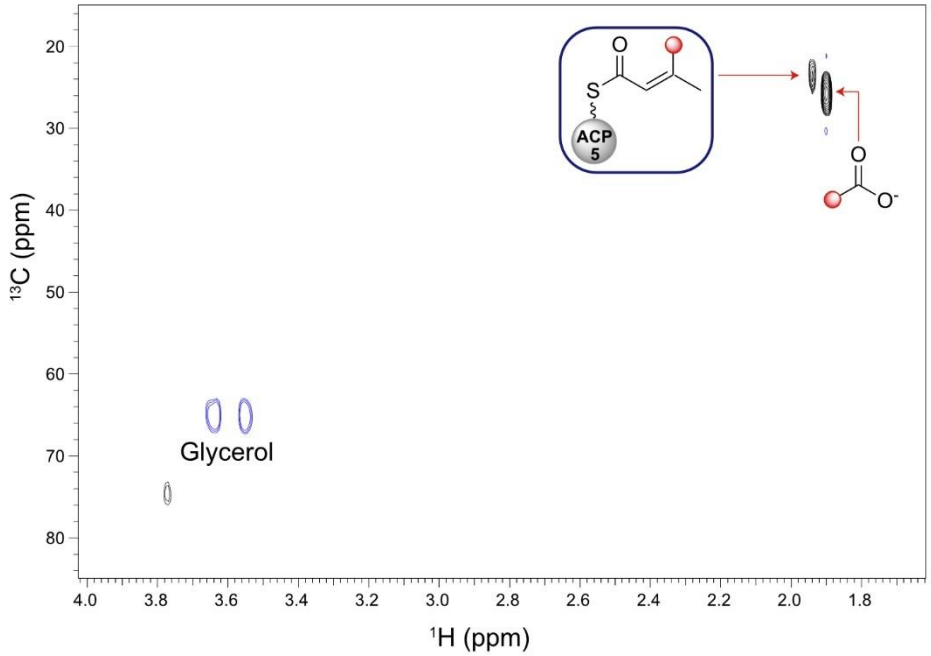


Figure 57 ^{1}H-^{13}C HSQC spectrum of the reaction Acac-ACP5 + [^{13}C]-Ac-BatA + BatC + BatD + BatE. Correlations for the [^{13}C]-*endo*-β-methyl product (23.49 ppm/1.95 ppm) and [^{13}C]-acetate (26.00 ppm/ 1.91 ppm) were observed and these were consistent with the chemical shifts for ACP4.

2.2.5 ^{13}C NMR assay for 4M

Having demonstrated the power of the methodology, the next step was to apply it to the di-domain 4M to determine the function of the mECH domain. Working with 4M was likely to present more challenges than ACP4 alone. Firstly, the protein is 43 kDa and dimerises to an 86 kDa complex, which is harder to analyse by NMR. Secondly, the influence of the mECH domain on the loading of the ACP and the relative stability of intermediates was unknown. Finally, the larger protein mass would make it harder to achieve high concentrations of ACP-bound intermediates, affecting the sensitivity of the assay. It would also require a large mass of protein per experiment, therefore requiring good protein expression and handling to prevent sample loss.

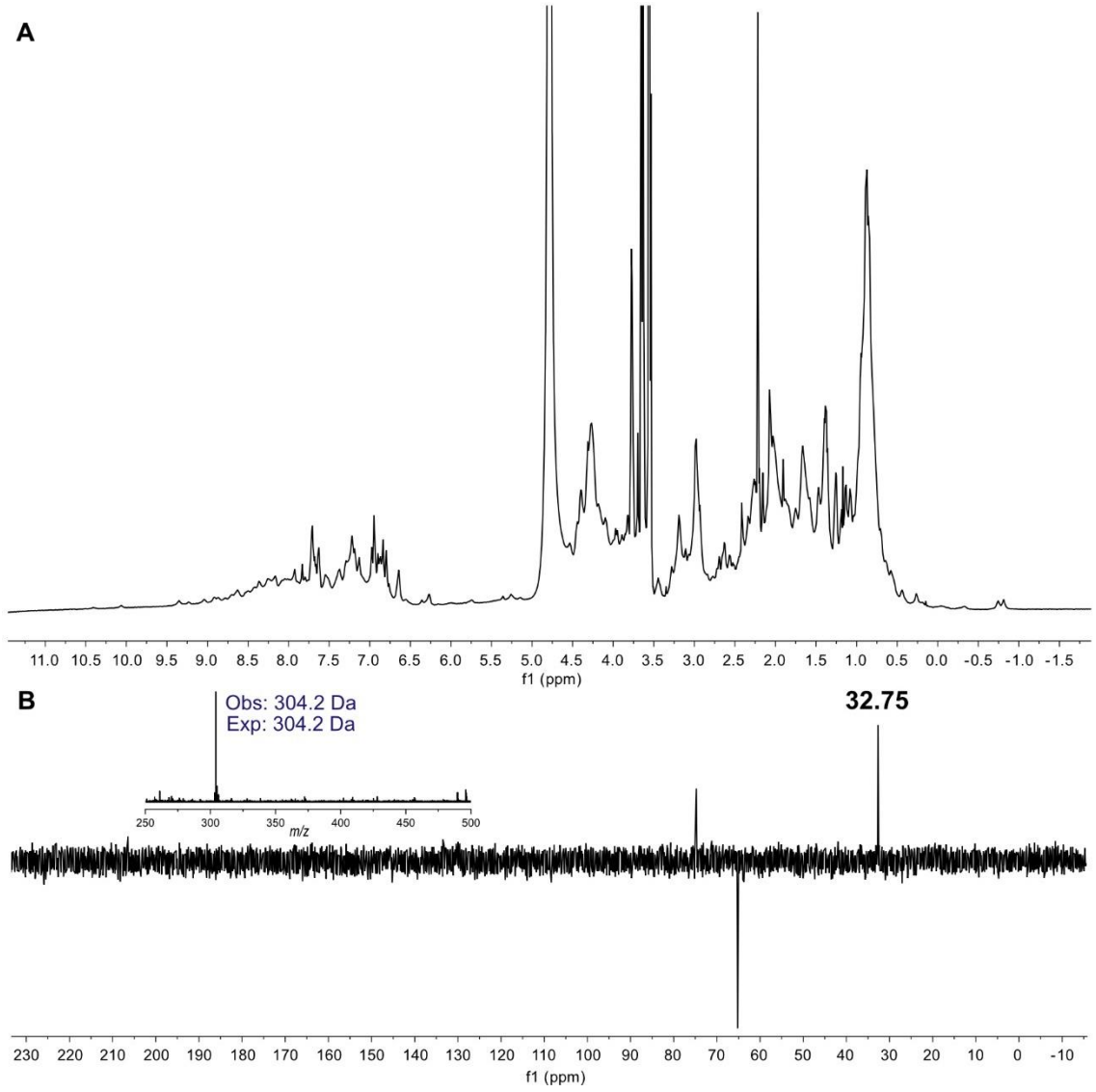


Figure 58 A) ^{1}H spectrum of [^{13}C]-Ac-4M. B) ^{13}C DEPT spectrum of [^{13}C]-Ac-4M gave a peak at 32.75 ppm along with MS analysis by Ppant ejection (304.2 Da). The peaks at 65.29ppm and 74.87 ppm correspond to glycerol.

[^{13}C]-Ac-pantetheine **85** was loaded onto 4M, desalted into NMR buffer and concentrated to 150 µM. A ^{1}H spectrum was recorded and, as expected, weaker and broader peaks were observed (Figure 58A). However, upfield shifted methyl peaks that could be correlated to ACP4 were present, indicating the ACP is folded. A DEPT spectrum was then recorded and the characteristic positive CH_3 peak at 32.75 ppm was observed (Figure 58B). This result meant that the mobility of the substrate attached to ACP4 was not influenced by the adjacent mECH domain and that intermediates were likely to be solvent exposed and therefore accessible to other catalytic enzymes and would be observable by ^{13}C NMR.

To reconstitute the β-branching pathway, Acac-pantetheine **62** was loaded onto apo-4M and after 1 h full loading was achieved. [^{13}C]-Ac-BatA and BatC were then added and the sample monitored by DEPT NMR, showing the conversion to [^{13}C]-HMG-4M (Figure 59). Once again, a significant proportion of non-productive hydrolysis to form [^{13}C]-acetate was also observed. Brief optimisation of this step found that a lower concentration of BatC reduced the production of [^{13}C]-acetate. Hydrolysis could also be minimised by the stepwise addition of the desired amount of BatC and [^{13}C]-Ac-BatA in two portions, with the second addition 30 mins after the first. Whilst not preventing hydrolysis, a greater proportion of [^{13}C]-HMG-4M was obtained.

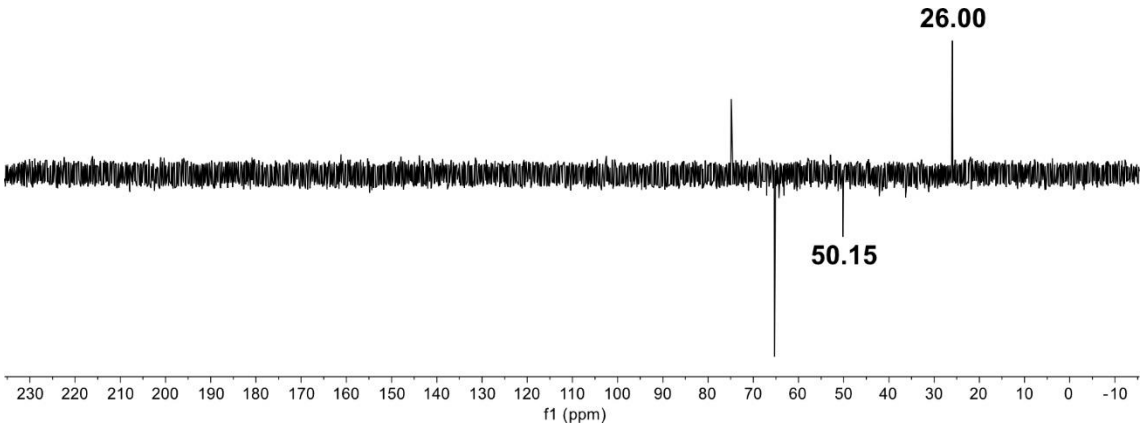


Figure 59 ^{13}C DEPT spectrum showing the formation of [^{13}C]-HMG-4M. It was not possible to analyse this intermediate by MS.

Freshly formed [^{13}C]-HMG-4M was incubated with BatD to determine whether mECH acts as a decarboxylase in the absence of BatE. The sample was monitored by DEPT NMR and pleasingly a new, negatively phased peak at 118.63 ppm was observed (Figure 60C). MS analysis of the sample gave a Ppant ejection ion at 344.2 Da, consistent with a [^{13}C]-unsaturated β-branch. Previous, analogous assays containing ACP4 did not show a peak at 344.2 Da by Ppant ejection that might arise from spontaneous

decarboxylation. Additionally, no *endo*-β-methyl peak was observed around 23.51 ppm, in accord with the new peak at 118.63 ppm being due to the exo double bond. A second peak at 42.10 ppm, previously only identified in the HSQC of the [^{13}C]-HMG-ACP4 and BatD reaction, was also present. As previously discussed, it is likely that this peak corresponds to [^{13}C]-MG-4M.

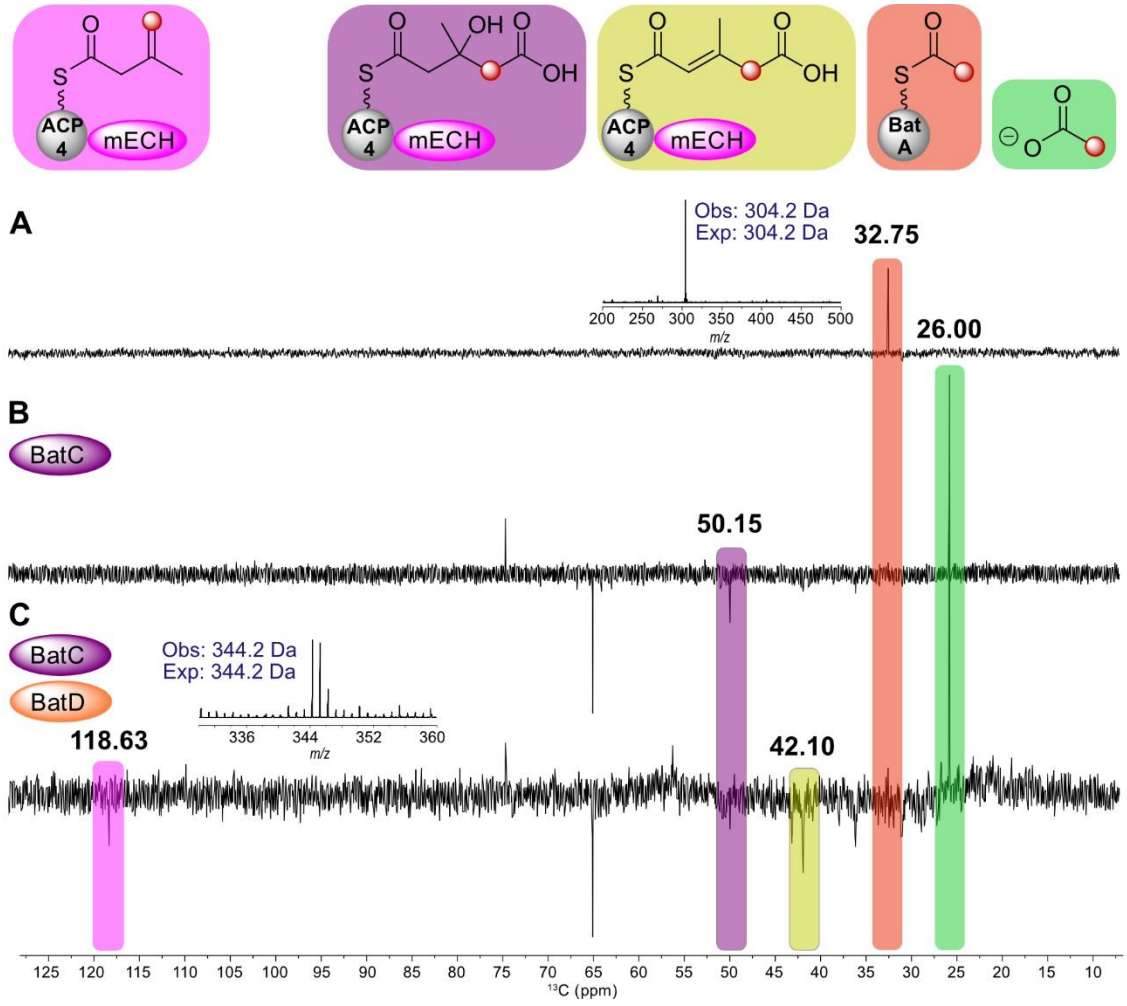


Figure 60 Stack of ^{13}C DEPT-edited spectra and MS analysis by Ppant ejection assay of the β-branching pathway for the di-domain 4M to give an *exo*-β-methyl branch. A) [^{13}C]-Ac-BatA, B) Acac-4M + [^{13}C]-Ac-BatA + BatC, C) Acac-4M + [^{13}C]-Ac-BatA + BatC + BatD.

To validate this result, β-*exo*-pantetheine **81** was prepared and analysed by NMR (work carried out by Angus Weir). Due to the propensity of the double bond to migrate into conjugation with the thioester, it was only possible to prepare a mixture of enoyl-pantetheine **64** and β-*exo*-pantetheine **81** (Figure 61A). The ^{13}C NMR of β-*exo*-pantetheine **81** contained a peak at 118.36 ppm consistent with the chemical shift of the *exo*-β-methyl formed on 4M (Figure 61B). The slight chemical shift perturbation arises from the attachment of the intermediate to an ACP resulting in a subtly different chemical

environment relative to the pantetheine. This result proves that the mECH domain acts as a decarboxylase which produces an *exo*-β-methyl by C-2 reprotonation (Figure 61C). This reactivity also rules out the possibility that mECH is an isomerase, as the *exo*-β-methyl is formed directly and does not require BatE.

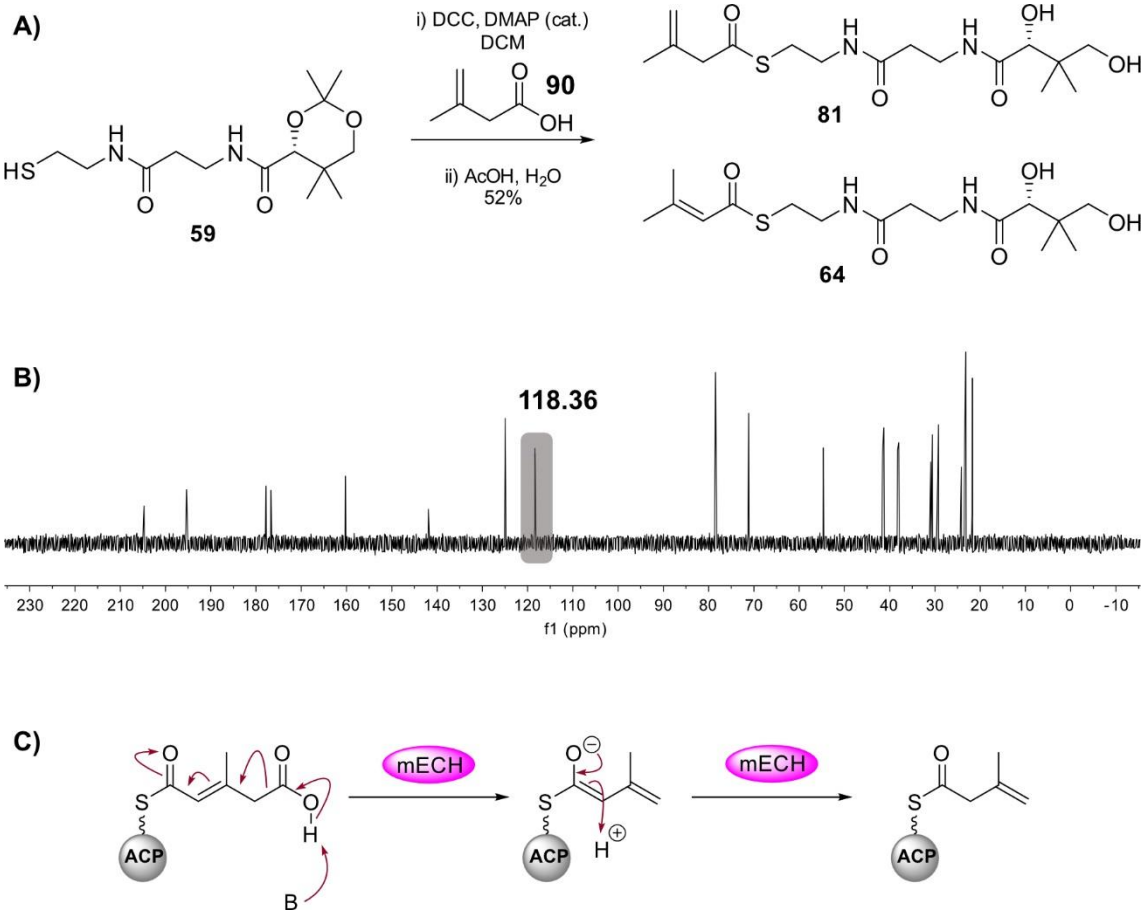


Figure 61 A) Synthesis of β-*exo*-pantetheine **81** as a mixture with enoyl-pantetheine **64**. B) ^{13}C spectrum of β-*exo*-pantetheine **81** contains a peak for the *exo*-methylene at 118.36 ppm (grey). C) Mechanism of decarboxylation and C-2 reprotonation by mECH.

Finally, [^{13}C]-HMG-4M was incubated with both BatD and BatE. The reaction contains two decarboxylase domains, creating a competition for the formation of either an *endo*- or *exo*-β-methyl. DEPT spectra showed the formation of a peak at 118.63 ppm and 23.51 ppm, which correlate to an *exo*-β-methyl and an *endo*-β-methyl respectively (Figure 62). It was not possible to validate this result by MS due to low forward flux of the pathway resulting in low product formation.

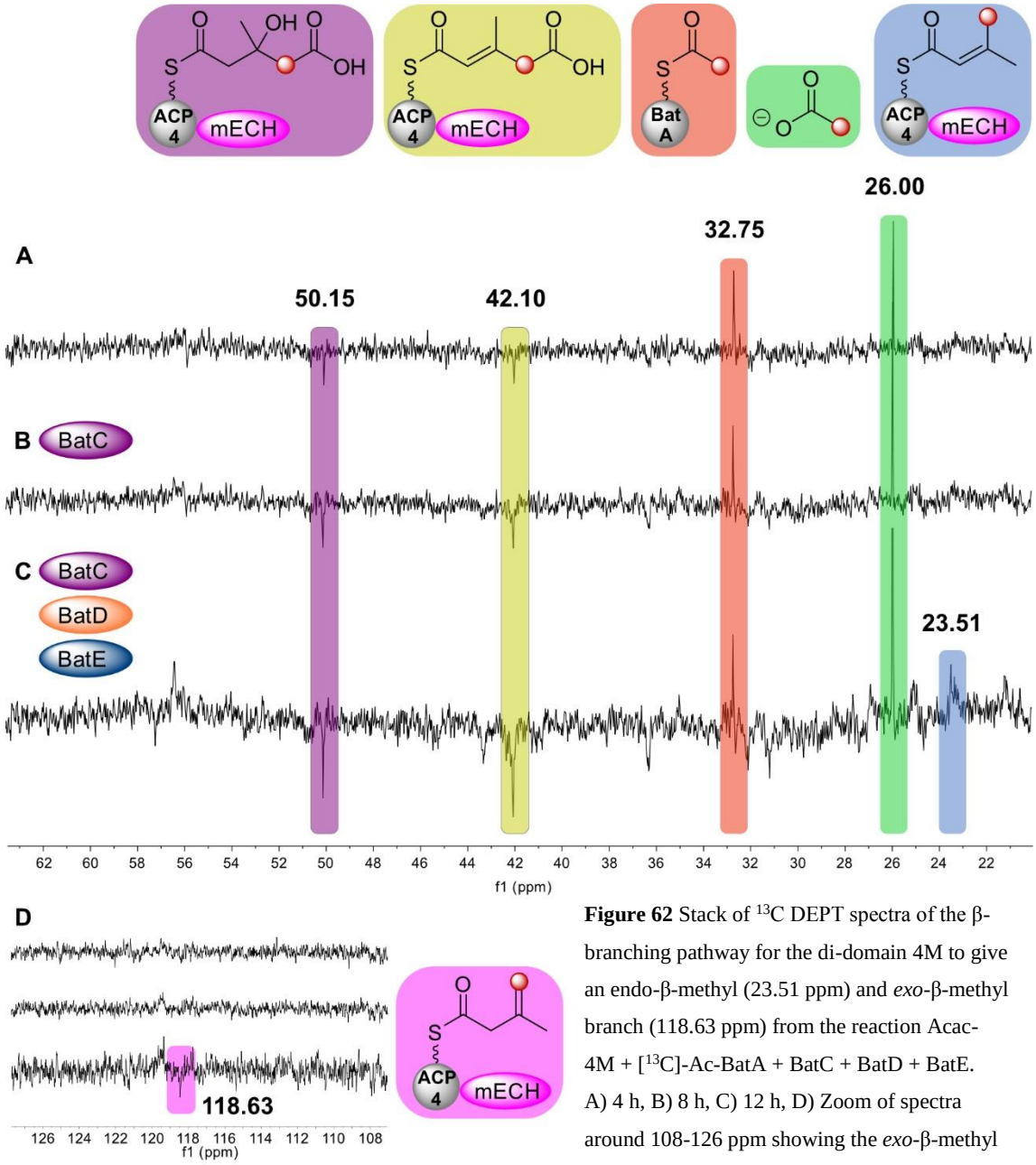


Figure 62 Stack of ^{13}C DEPT spectra of the β-branching pathway for the di-domain 4M to give an endo-β-methyl (23.51 ppm) and *exo*-β-methyl branch (118.63 ppm) from the reaction Acac-4M + [^{13}C]-Ac-BatA + BatC + BatD + BatE. A) 4 h, B) 8 h, C) 12 h, D) Zoom of spectra around 108-126 ppm showing the *exo*-β-methyl forming after 12 h. The persistence of the dehydrated intermediate at 42.10 ppm suggests poor efficiency of the reaction.

The presence of [^{13}C]-Ac-BatA and [^{13}C]-HMG-4M, in addition to the dehydrated intermediate at 42.10 ppm suggested that the flux through the pathway was inefficient or stalled. This points to the limitation of the model system used to reconstitute the β-branching pathway *in vitro*. The 3D architecture of the modular PKS, correct biosynthetic intermediate and directionality of the pathway that offloads substrates from the ACP will have a major impact on the flux of the system which may lack some components required for tight control.

2.2.6 Bioinformatic analysis of ECH_2 domains

The differential formation of *endo*-β-methyl and *exo*-β-methyl branches *via* an ECH_2-catalysed decarboxylation/selective reprotonation observed for BatE and mECH respectively suggests a key difference in active site architecture to ensure correct β-branch formation. Sequences of ECH_2 domains that produce either *endo*-β-methyl and *exo*-β-methyl branches were aligned using Clustal Omega (Figure 63)[143]. All domains are part of *trans*-AT PKS pathways, except for the well-studied *cis*-AT polyketides curacin and jamaicamide.

Figure 63 Truncated sequence alignment for ECH_2 domains producing *endo*- and *exo*-β-methyl branches. The conserved catalytic histidine and OAH-forming residues are shown in red and orange respectively. A PGxG motif that differentiates the *endo*- and *exo*-β-methyl producing domains is shown in green along with a key aromatic residue (Phe or Tyr) in blue. A) ECH_2 domains from *trans*-AT PKS that produce an *endo*-β-methyl branch contain a PGMG motif and conserved phenylalanine (Phe132, BatE). B) ECH_2 domains from *cis*-AT PKS that produce *endo*- (CurF) and *exo*- (JamJ) β-methyl branches in which the PGxG motif is not observed. C) ECH_2 domains from *trans*-AT PKS that produce *exo*-β-methyl branches contain a PGAG motif and a conserved tyrosine (Tyr129, Bat3_mECH).

Using the literature structure of CurF ECH_2, the catalytic His240 (CurF) and OAH-forming residues Ala78 and Gly118 (CurF) were conserved across all domains.[55] Careful analysis of the alignment revealed a PGMG (139-142, BatE) motif present in the *endo*-β-methyl producing and an equivalent PGAG (136-139, mECH) motif in the *exo*-β-methyl producing ECH_2 domains. The only exceptions to the motif were CurF and JamJ ECH_2 which contain PVGA and PVGS motifs respectively. The absence of the conserved motif in these *cis*-AT PKS domains might arise from the separate evolution of these pathways from the *trans*-AT PKS.[29] Another difference in the two sets of ECH_2 domains was an aromatic amino acid 7-residues upstream of the PGxG motif. The *endo*-β-methyl

producers have a phenylalanine (Phe132, BatE) whilst the *exo*-β-methyl producers contain a conserved tyrosine (Tyr129, mECH).

To explore the 3D architecture of the ECH_2 domains and the position of the key residues, homology models of mECH and BatE from the kalimantacin pathway were constructed using the Phyre2 and I-Tasser servers. The structures generated by each of the servers displayed the same overall architecture, consisting of the classic CS fold, but there was a dramatic difference in the position of the C-terminal domain. These α-helical domains are vital in the formation of oligomers in the CS, with trimers or dimers of trimers the most common oligomeric state observed.[100] Depending on the position of the C-terminal helix, the interactions may be categorised as monomer self-association, intra-trimer or inter-trimer. In the monomer self-association, the C-terminal helix forms the active site of the monomer by folding back across the core structure.[144] However, by orientating the helix away from the core, monomer structure, it may form the active site of an adjacent monomer within the trimeric structure (intra-trimer) or across the trimer-trimer interface (inter-trimer).[145]

The models generated by Phyre2 contained the self-association fold with the active site created by the C-terminal domain of the monomer. In contrast, both I-Tasser models contained a C-terminal domain folded away from the monomer structure, as expected for an inter/intra-trimer oligomer. The CurF ECH_2 domain was shown to form the self-association fold in the solid-state trimer solved by Smith *et al*.[55] Therefore, the Phyre2 models were used to map the key catalytic residues in the active site.

His231 (BatE), His 228 (mECH) and the OAH-forming residues - Gly67/Gly109 in BatE and Gly64/Ala106 in mECH - were mapped onto their respective structures and showed very similar positions around the active site (Figure 64A and D). The C-terminal domain encloses the active site with the final helix creating a capping loop over the entrance to the substrate channel. The PGxG motif is located in a loop adjacent to the active site in both BatE and mECH. In BatE, Met141 sits at the bottom of the active site, close to the OAH-forming Gly67 and Gly109, creating a hydrophobic surface of the narrow channel that leads to the catalytic His231 and forcing Phe132 away from the active site (Figure 65 B and C).

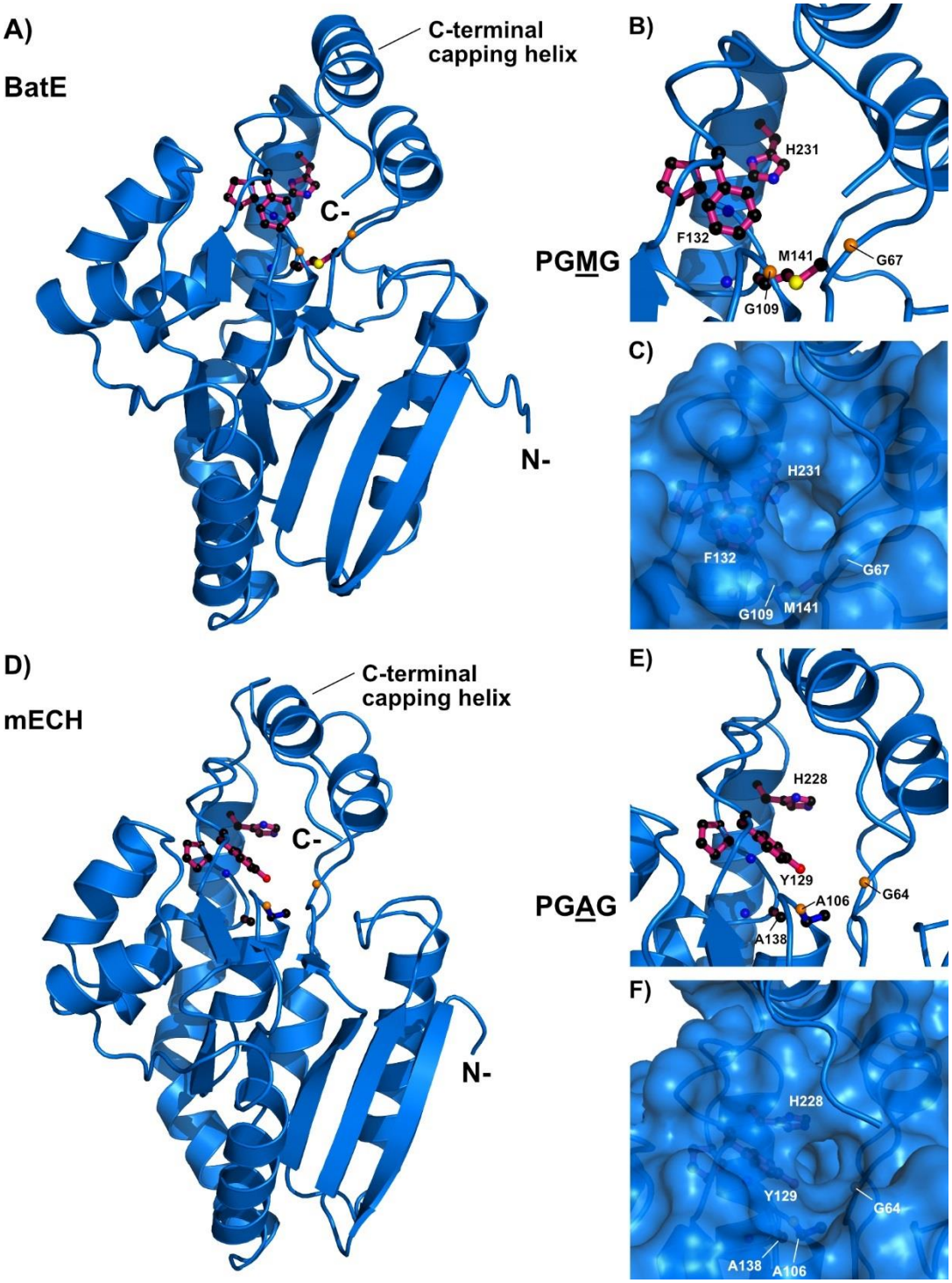


Figure 64 Phyre 2 homology models for BatE and mECH both represented as the monomer self-association fold. A) Homology model of BatE. B) Zoom of BatE active site with the catalytic H231 and OAH-forming residues G67 and G109 (orange). M141 from the PGMG motif blocks the bottom of the active site and along with F132 forms a hydrophobic compartment. C) Surface representation of the active site shows a channel for substrate feeding lined by the key catalytic residues. D) Homology model of mECH. E) Zoom of the mECH active site with the catalytic H228 and OAH-forming residues G64 and A106 (orange) in conserved positions. A138 from the PGAG motif does not block the bottom of the active site and Y129 is positioned close to the OAH and is a likely proton-source for the C-2 reprotonation to form an *exo*-β-methyl branch.

In contrast, the equivalent Ala138 residue of the PGAG motif in mECH occupies a similar position to Met141, but the smaller alanine creates a more open cavity (Figure 64 E and F). This allows Tyr129 to occupy a position within the active site cavity, proximal to the OAH-forming residues. It is therefore hypothesised that Tyr129 is the proton source that, following decarboxylation, reprotonates at the C-2 position to form an *exo*-β-methyl branch. Due to the absence of a polar residue in the BatE active site, it is proposed that His231 may deprotonate and, following decarboxylation, reprotonate the C-4 position to give the *endo*-β-methyl branch. The methionine residue may facilitate this reaction by forming the narrow cavity to orientate the substrate towards His231 compared to the more open cavity arising from Ala138 in mECH.

2.2.7 Mechanisms for β-branch selectivity

The difference in ECH_2 sequence and structure offers a hypothesis for the formation of both the *endo*- and *exo*-β-methyl branches that were observed experimentally. However, it is also important to rationalise how the kalimantacin biosynthetic pathway – containing several enzymatic β-branching pathways - achieves the required selectivity to produce a single polyketide product. This may be understood by considering the type I PKS oligomerisation, 3D architecture and the interaction of *trans*-acting catalytic domains.

The ECH_2 integrated into CurF is the only ECH_2 domain in the curacin biosynthetic pathway and the only example of a β-branching ECH_2 structurally characterised in the literature.[55] It was shown to be trimeric in the solid state, but a concentration-dependant equilibrium between trimeric and lower oligomeric species was observed in solution. Analysis of the monomeric units showed that they form the self-association fold, whereby the active site is covered by the C-terminal domain of the same monomer. This would allow for a fully functional enzyme without the requirement for oligomerisation. The authors state that it is currently not clear how the modular, trimeric ECH_2 domain might associate with downstream domains (KS, DH) that are known to dimerise.

In the kalimantacin biosynthetic pathway, the free-standing ECH_2 (BatE) was shown to be trimeric in solution whereas the 4M di-domain was dimeric. Whilst the *trans*-acting BatE oligomerises as would be expected of a CS protein, the dimeric 4M construct creates uncertainty over the assembly of type I PKS modules that contain both, normally trimeric, ECH_2 domains as well as, normally dimeric, chain extension (KS) domains. Either the ECH_2 adopts a less common dimeric structure or trimerisation at the ECH_2 domain

occurs, creating large, complex oligomeric structures. The formation of biosynthetic megacomplexes was reported by Keatinge-Clay *et al.* who identified a motif capable of mediating this association.[146] How these domains ultimately associate will undoubtedly affect the possible interactions of the modular ACPs that are required to deliver the growing polyketide chain *in-cis* as well as interact with *trans*-acting domains.

In the formation of the C-7 *exo*-β-methyl branch at ACP4, BatD (trimeric in solution) dehydrates HMG-ACP4 prior to decarboxylation. The in-*cis* mECH then decarboxylates the dehydrated intermediate and selectively reprotonates it to form the *exo*-β-methyl branch. Our *in vitro* results show that BatE can interact with ACP4 expressed as a single domain and as part of the 4M di-domain to form an *endo*-β-methyl branch. Additionally, the *in vivo* formation of a C-7 sat.-β-methyl branch in a kalimantacin minor metabolite **69** suggests it is possible for BatE (and subsequently BatK) to interact with ACP4 in the wild-type system. However, the pathway shows high efficiency in the formation of the *exo*-β-methyl branch due to the low titres of the minor compound. It is hypothesised that this high selectivity is achieved kinetically, with faster flux of biosynthetic intermediates at ACP4 in-*cis* than in-*trans*. mECH would therefore act faster on the dehydrated intermediate than BatE, resulting in *exo*-β-methyl formation. If BatE were to act faster than mECH, the result would be the formation of the minor sat.-β-methyl branch *via* BatK reduction.

An alternative mechanism for selectivity is the steric occlusion of BatE from ACP4. This would require ACP4 to discriminate between the two trimeric proteins BatD and BatE. Following BatD-catalysed dehydration, a structural rearrangement would have to take place to prevent BatE from decarboxylating the dehydrated intermediate. Structural rearrangements of a whole module from the modular *cis*-AT PKS that produces pikromycin were presented by Skiniotis *et al.* in 2014.[34] Using cryo-EM, the authors were able to capture a series of distinct catalytic states and demonstrate the rearrangement required between each step. Each structural rearrangement is necessary to pass the ACP-bound intermediate from one catalytic domain to the next to produce the correct polyketide. As each domain is fused as part of a single polypeptide chain, the position of all domains is well defined for each catalytic step.

In the kalimantacin pathway, however, the catalytic domains are acting in-*trans*. This means that BatD must associate with the modular ACP rather than simply rearranging to

a defined location to deliver the substrate as happens for in-*cis* substrate channelling. Following dehydration, a structural rearrangement would have to occur quickly to prevent BatE association with the ACP. The timing of this structural rearrangement and recognition of each catalytic domain must be very strict to prevent the wrong substrate being delivered to each of the in-*trans* catalytic domains. It may therefore be a substrate-controlled process, in which the ACP must recognise pre- and post- catalysis substrates. These steric arguments, whilst plausible, would need to be rigorous to prevent aberrant β-branch formation, and perfect fidelity is not achieved as BatE/BatK are able to produce the sat.-β-methyl minor metabolite **69**.

Whilst the CS has been extensively studied, there is little structural information available for domains that interact with ACP-bound rather than CoA-bound substrates, particularly the ECH domains from HCS cassettes. Understanding their structures, oligomerisation in a type I PKS and active site residues will all be critically important to understand how the channelling of substrates occurs in β-branching modules. In particular, the control required for the selective interaction of modular ACPs with *trans*-acting domains is poorly understood and offers an exciting area for future research.

2.3 Kalimantacin Isolation, Analysis and Synthetic Studies

2.3.1 Natural product isolation and stability

Kalimantacin A shows promise as an antimicrobial agent, however, there was uncertainty over the stability of the molecule (Figure 65). In such a highly functionalised compound, the *E,Z*-diene may isomerise and cleavage of the chain *via* a retro-aldol reaction may occur. In addition to the natural diene geometry, small amounts of the *E, E*-isomer had been isolated from *P. fluorescens* by previous members of the Willis group.[97]

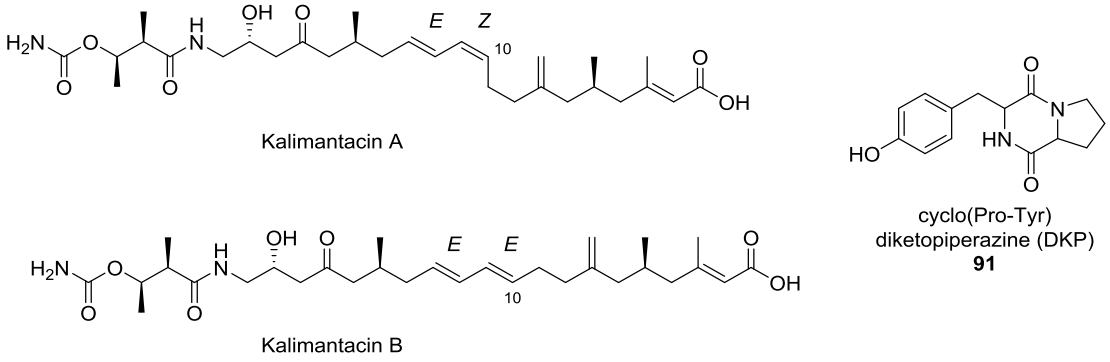


Figure 65 Structures of kalimantacin A and B and the co-eluting side DKP side product **91**.

To investigate this further, two separate stocks of *P. fluorescens* were streaked onto LB agar, a pre-culture prepared and used to inoculate a 1 L culture. After 4 days, the cells were separated, and the supernatant extracted with ethyl acetate to give a crude oil. By LC/MS, both fermentations had produced kalimantacin and the separate crude extracts were purified by column chromatography. After the first column (MeOH/CH_2Cl_2/AcOH), kalimantacin A was mixed with diketopiperazine **91** (DKP) as previously reported.[97] This material was further purified using MeOH/EtOAc/AcOH as the eluent. Batch 1 gave 95% pure kalimantacin which was characterised by ^{1}H, ^{13}C NMR and HRMS and the data recorded were in good agreement with the literature (Figure 66A).[78]

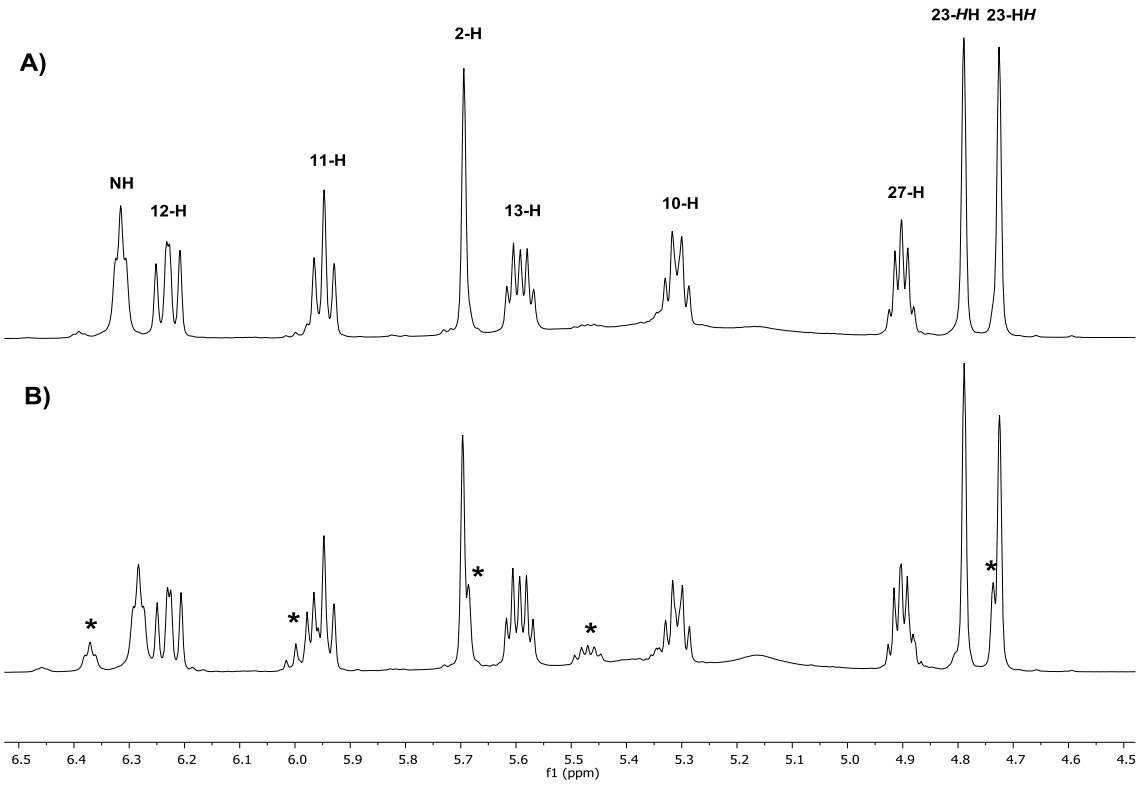


Figure 66 Expansion of ^{1}H NMR ($CDCl_3$) spectra of the kalimantacins. A) The compound from batch 1 was in agreement with the literature assignment of kalimantacin A containng the *E,Z*-diene .[78] B) In batch 2, kalimantacin A was isolated with ~80% purity, however, a small amount of kalimantacin B containing the *E,E*-diene was co-purified.[76] The peak at 5.47 ppm corresponds to the 10-H of kalimantacin B, whilst other minor peaks were also evident (*****).

In addition to kalimantacin A, the ^{1}H NMR spectrum of batch 2 showed several minor peaks that were assigned to the *E,E*-diene moiety, with the multiplet at 5.47 ppm (10-H) being one of the most characteristic (Figure 66B). This was assigned from the ^{1}H NMR data published for kalimantacin B produced by the *alcaligenes* species.[76] This double bond isomer has previously been observed both in the isolation of kalimantacin-like natural products, but also in the chemical synthesis of related compounds.[98] One

hypothesis for the formation of both isomers was that the acidic work-up conditions catalysed isomerism of the *E,Z*-diene to the *E,E*-diene. To test this theory, the fermentation was repeated, and half the supernatant was acidified to pH 2 with HCl then extracted, whilst the remaining half was extracted at pH 7.

Following the same chromatography procedure, the yields from the non-acidified (18 mg) and acidified (19 mg) work ups were comparable and in both cases a small quantity of the *E,E*-isomer (10%) could be observed as shown in Figure 67. This result suggests that there is not an acid-driven isomerisation between the *E,Z* natural product and the *E,E* minor isomer. However, it does demonstrate the inconsistency in the production of the minor isomer as a different ratio was obtained in each of the fermentations.

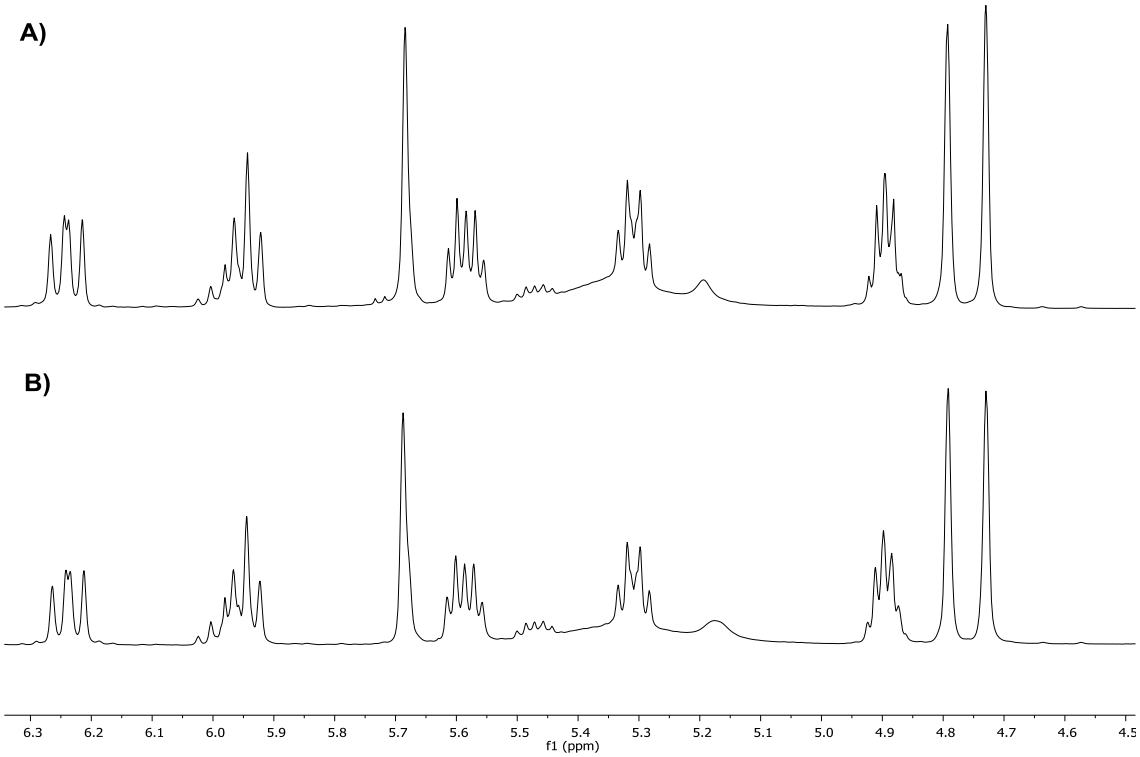


Figure 67 Expansion of ^{1}H NMR ($CDCl_3$) spectra for kalimantacin. A) Isolated under normal conditions at pH 7 and B) following acidification to pH 2 with HCl. Around 10% kalimantacin B is present in both samples, determined by the characteristic 10-H peak at 5.47 ppm.

To understand more about the stability of kalimantacin A, 2 mg/ mL of kalimantacin from both batch 1 and 2 were stored separately in an NMR tube in $CDCl_3$ under normal lab conditions of 25 °C and exposed to light. ^{1}H NMR spectra were recorded every 7 days for 5 weeks (Figures 68 and 69).

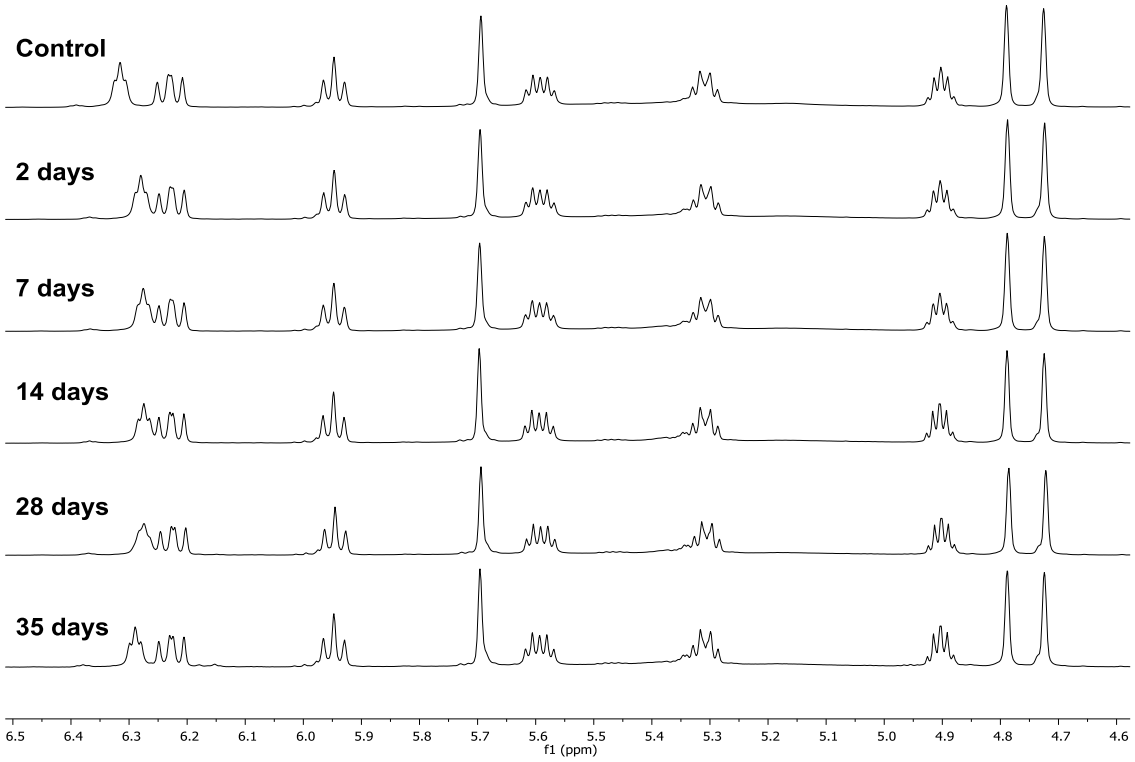


Figure 68 Stack of expanded ^{1}H NMR spectra for kalimantacin A isolated in batch 1 and stored in $CDCl_3$ at 25 °C.

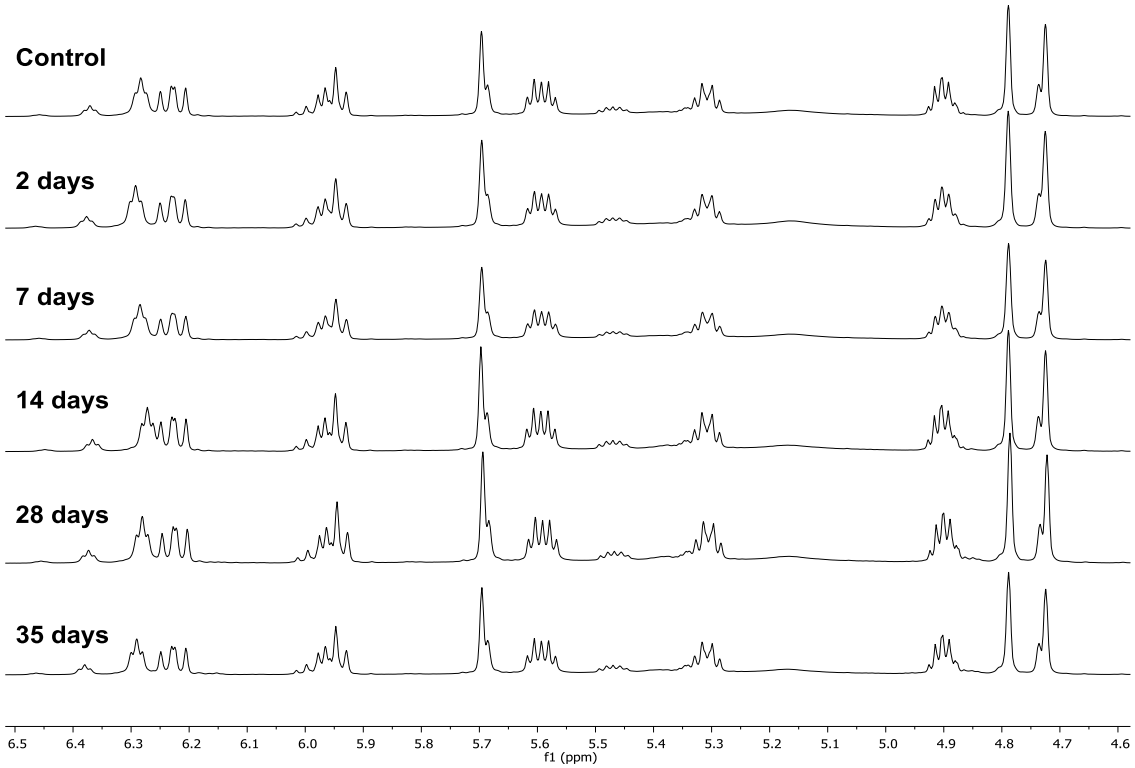


Figure 69 Stack of expanded ^{1}H NMR spectra for kalimantacin A/B isolated in batch 2 and stored in $CDCl_3$ at 25 °C.

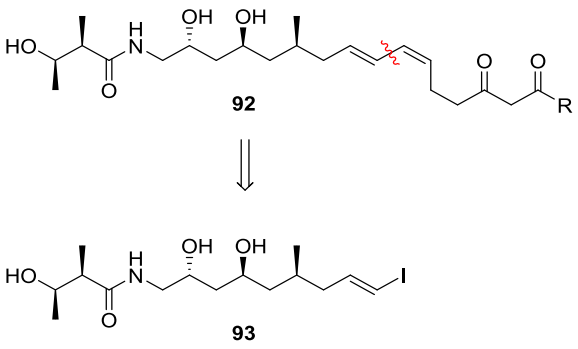
During the time course, no change in the proportion of *E,E*- and *E,Z*- isomers was observed, indicating the stability of the molecule to light, ambient temperature and the mildly acidic conditions provided by the deuterated chloroform. This result suggests that the proportion of *E,Z*- and *E,E*- isomers is determined during the fermentation and isolation procedure and remains constant thereafter.

Further work conducted by Angus Weir (a PhD student in our group) has resulted in fermentation broths that give rise to 60% kalimantacin A and 40% of the *E,E*-isomer kalimantacin B. Furthermore, excluding light during fermentation yielded variable ratios of diene geometry and storing a pure sample of kalimantacin A at room temperature resulted in isomerism to kalimantacin B. Despite extensive work, it is still not possible to completely control the proportion of diene isomers isolated from fermentation broths. Conflicting evidence about the isomerism of the *E,Z*-diene to an *E,E*-diene in dilute solution and neat suggest that the compound is unstable around this moiety. A better understanding of this process is required.

2.3.2 Synthetic studies

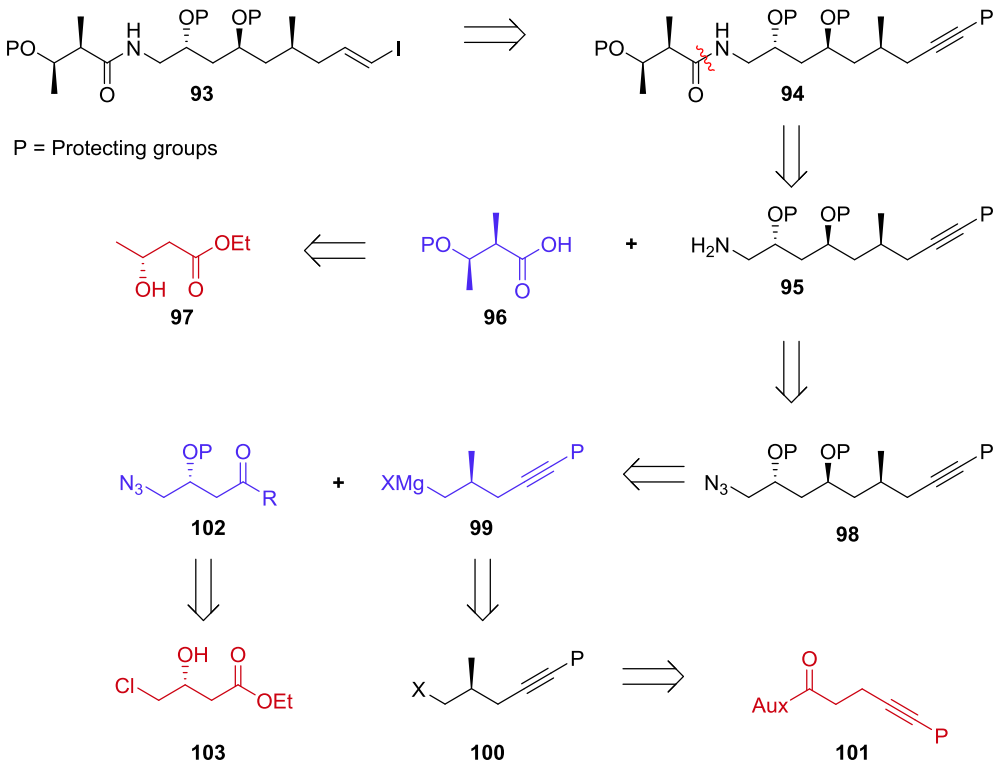
To study a complex pathway like the kalimantacin PKS-NRPS, a reliable source of material either from natural sources or chemical synthesis is required. Natural product isolation from the WT and mutant strains provides a rich source of late-stage compounds.[96] However, early stage intermediates and chemical probes of the biosynthetic pathway are also necessary. Most commonly, simplified structures that contain several key functional groups for enzymatic transformation are included due to the synthetic accessibility of these simplified intermediates.[39] This has been demonstrated in Chapter 2.1-2.2, however, use of the correct substrate may increase efficiency or specificity *in vitro*. Preliminary work was undertaken to synthesise a key building block to provide kalimantacin-like biosynthetic intermediates.

Diene **92** is the proposed biosynthetic intermediate attached to ACP3/4 which, following interaction with the HCS cassette, yields an *exo*-β-methyl branch. Retrosynthesis of **92** through disconnection of the diene moiety would lead to a key vinyl iodide fragment **93** (Scheme 28), which could be used in the synthesis of β-branching biosynthetic intermediates, as well as in the total synthesis of kalimantacin or the diol precursor. Furthermore, access to analogues *via* a total synthesis could be used in SAR studies to determine the key functionalities in the active compound.



Scheme 28 Target biosynthetic intermediate **92** where R is a thioester linkage to CoA/ACP.

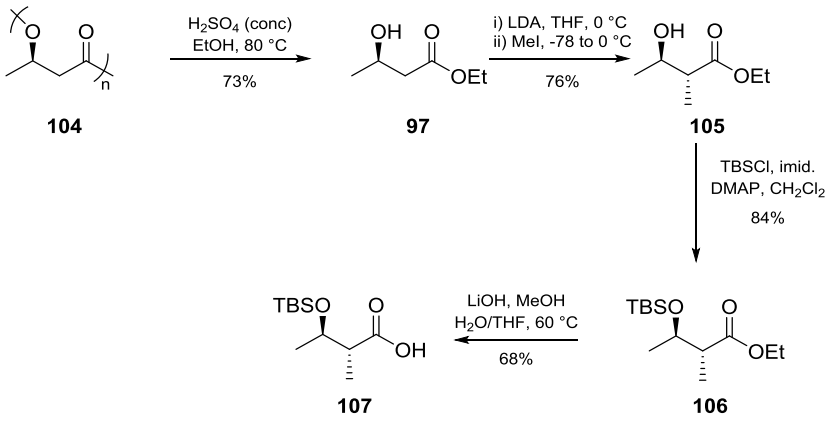
The synthetic efforts of Dr Iain Thistlethwaite and Dr Freya Bull led to the elucidation of the relative and absolute stereochemistry of kalimantacin A.[96] Utilising the knowledge gained during that work, a synthetic route to key vinyl iodide **93** was designed. The vinyl iodide originates from a disconnection through the diene system of kalimantacin diol which may be constructed in the forward synthesis *via* a metal-catalysed sp^2-sp^2 cross coupling.



Scheme 29 Retrosynthetic analysis of key vinyl iodide fragment **93** leading to readily available or easily synthesised starting materials (shown in red). The key compounds for synthesis are carboxylic acid **96**, Grignard reagent **99** and azide **102** which are shown in blue.

A retrosynthetic analysis of vinyl iodide **93** is shown in Scheme 29. The vinyl iodide would be prepared from a protected alkyne **94**. Disconnection of the amide bond furnishes a protected acid **96** which can be made from readily available ethyl (*R*)-3-hydroxybutyrate **97**. Amine **95** would be masked as azide **98**, allowing for the addition of Grignard reagent **99** into the carbonyl compound **102** which may either be the aldehyde or Weinreb amide. This in turn can be prepared from readily available ester **103**. Finally, the Grignard reagent **99** can be synthesised from halide **100** *via* an asymmetric alkylation of **101**.

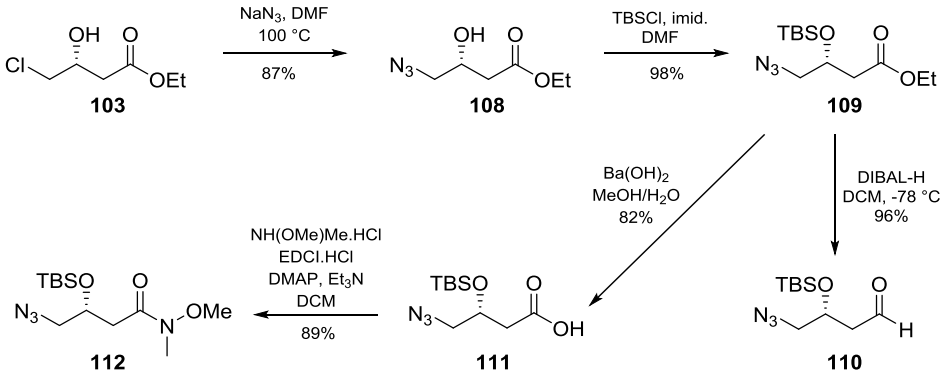
Adapting the previously reported route,[97] poly-(*R*)-3-hydroxybutanoate **104** was refluxed for 7 days with EtOH and H_2SO_4 (Scheme 30). The reaction was found to proceed smoothly without the addition of dichloroethane as a co-solvent and, after purification by vacuum distillation, ester **97** was isolated in a 73% yield, comparable to the previous report.[97] A Fráter-Seebach stereoselective alkylation of **97** with LDA and MeI afforded **105** as a single diastereomer. TBS protection of the alcohol and ester hydrolysis under basic conditions furnished the desired acid **107** in 57% yield over the two steps.



Scheme 30 Synthesis of carboxylic acid **107** from poly-(*R*)-3-hydroxybutanoate **104**.

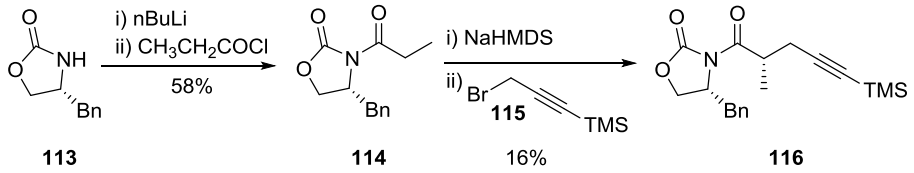
For the azide fragment, starting from commerically available alcohol **103**, the chloride was displaced in an S_N2 reaction with sodium azide (Scheme 31). The R_f of the starting material **103** and product **108** were the same and so the reaction could not be monitored by TLC analysis. Upon work-up, ^{1}H NMR and IR spectroscopy indicated the consumption of starting material and the formation of the azide as indicated by the strong asymmetric stretching band at 2098 cm^{-1}. Alcohol **108** was protected as silyl ester **109** in excellent yield. **109** was treated with DIBAL-H at -78 °C for 10 min and after work up and purification afforded aldehyde **110** in 96% yield. Hydrolysis of ester **109** under basic conditions also proceeded smoothly, furnishing acid **111** that was coupled with *N*,*O*-

dimethylhydroxylamine hydrochloride to give Weinreb amide **112**. Having both compounds (**110** and **112**) available allows flexibility in the later synthetic route. Grignard addition to the aldehyde **110** is likely to proceed with no stereocontrol of the resulting alcohol. It is expected that an oxidation/selective reduction would give the desired *trans*-relationship of the 1,3-diol. By reacting the Grignard **99** with Weinreb amide **112**, the expected ketone would be ready for selective reduction, omitting the need for oxidation. The azide functionality is required at this stage to prevent cyclisation to form an unwanted 6-membered ring as observed previously when using protected amines.[97]



Scheme 31 Synthesis of azide compounds containing either an aldehyde **110** or Weinreb amide **112** moiety.

With the azide fragments **110** and **112** in hand, synthesis of Grignard reagent **99** was required. The first approach to selective alkylation employed auxiliary **114** and TMS-protected propargyl bromide **115** as the alkylating agent (Scheme 32). However, the reaction was poor yielding (16%) and so the use of the opposite auxiliary with a methylating agent was investigated.

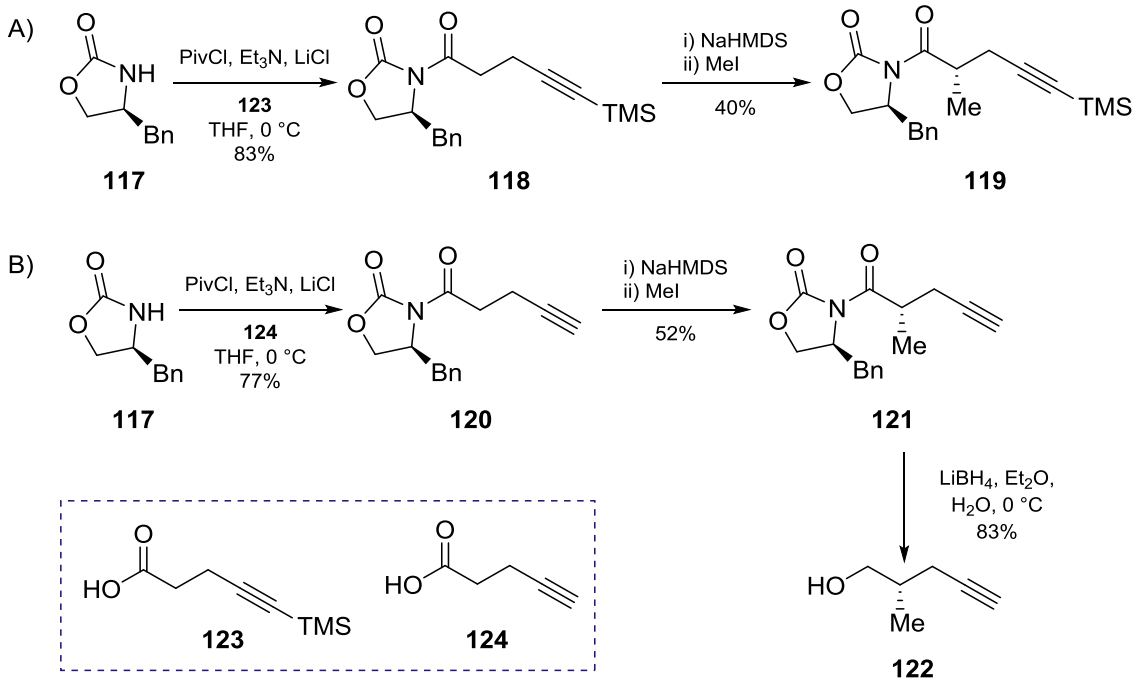


Scheme 32 Selective alkylation using Evan's auxiliary **114.**

Protected acid **123** was reacted with Evans auxiliary **117** *via* a mixed anhydride intermediate giving **118** in an 83% yield (Scheme 33). Despite using two or greater equivalents of base and TMSCl in excess, it was not possible to directly TMS protect pent-4-yn-oic acid **124**. As a result, protection of alcohol **125** and oxidation of the resulting compound **126** was required (Scheme 34). Acylated auxiliary **118** was treated

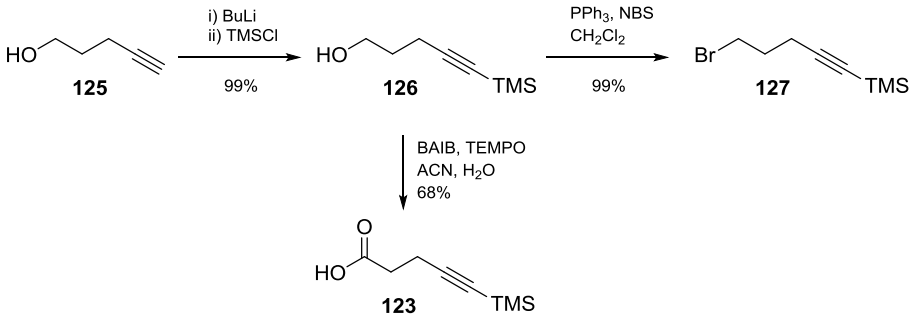
with NaHMDS at -78 °C then methyl iodide was added. The reaction was tried several times, including warming the reaction to around -50 °C to try to promote the alkylation, however, a maximum yield of 40% of **119** was obtained.

In parallel, the unprotected alkyne **124** was coupled to auxiliary **117** in an analogous process. Treatment of **120** with NaHMDS and MeI again furnished the desired product **121** but in a disappointing 52% yield. Interestingly given the similarities in pK_a of the α-proton and the alkyne proton, no alkylation of the terminal alkyne was observed. As a proof-of-concept, reductive cleavage of **121** gave the desired alcohol **122** in 83% yield.



Scheme 33 Synthesis and alkylation of Evan's auxiliaries using A) protected alkyne **118** B) unprotected alkyne **120.**

To explore the feasibility of using the desired Grignard reaction, the test substrate **127**, lacking the methyl group at C-2, was synthesised from **125** (Scheme 34). Converting **127** to the corresponding Grignard reagent was attempted employing conditions used within the group in the synthesis of a similar reagent.[98] Despite this, the desired Grignard was not formed, possibly due to the small scale of the reaction. Having synthesised fragments **107, 110** and **112**, the synthesis of Grignard reagent **99** and optimisation of the Evan's alkylation are the key steps in the construction of kalimantacin-like biosynthetic intermediates. However, time did not permit any further investigations.



Scheme 34 Synthesis of acid **123** and bromide **127**.

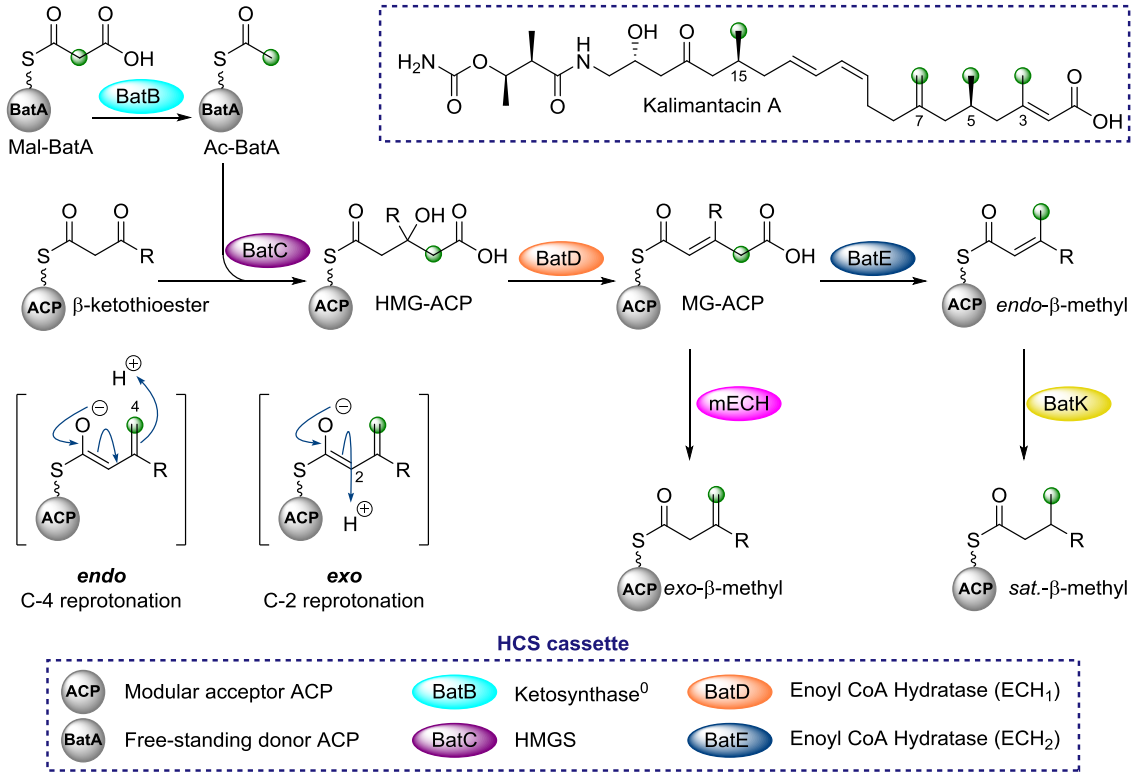
This initial work was built on by Angus Weir during his Master's research project and has subsequently been furthered by PhD student Jon Davies. The aim is to create a flexible synthetic route that allows for the synthesis of biosynthetic-intermediates, analogues of kalimantacin and to complete a total synthesis.

2.4 Conclusions and Future Work

2.4.1 Conclusions

Reconstitution of the kalimantacin β-branching pathway and monitoring by ^{13}C NMR and Ppant ejection assay has provided a powerful tool for the monitoring of complex, multi-enzyme catalysis of ACP-bound substrates. The simplicity of the single [^{13}C] label, sensitivity of high-field NMR spectrometers and complementation by MS has allowed for the detailed characterisation of each enzymatic step and robust identification of the intermediates formed (Scheme 35).

The HCS cassette, modular ACPs and tailoring enzymes from the kalimantacin gene cluster were cloned, expressed and purified. *In vitro* reconstitution of the β-branching pathway firstly demonstrated the formation of an *endo*-β-methyl branch. The α,β-unsaturation and stereochemistry of the dehydrated intermediate (MG-ACP) was confirmed by pantetheine- and ACP-bound NMR experiments. These techniques are more rigorous than those previously reported in the literature.[45] An ACP-bound *endo*-β-methyl was also shown to be a substrate for the reduction by BatK to form a saturated β-methyl branch. Finally, a previously unassigned mECH domain was identified in Bat3. The domain was excised from the PKS, expressed and purified. It showed no function in a range of MS and NMR experiments and structural characterisation revealed a mis-folded domain.



Scheme 35 Biosynthetic pathway for the formation of *endo*-β-methyl, *exo*-β-methyl and sat.-β-methyl branches by a single HCS cassette in kalimantacin biosynthesis.

This led to the expression of the di-domain ACP4-mECH (4M) and development of an NMR assay that utilised the specific incorporation of a single [^{13}C] label into key biosynthetic substrate mimics. By this NMR assay, the formation of an *exo*-β-methyl branch was demonstrated *in vitro* in the presence of BatD and 4M but the absence of BatE. This means mECH acts as a decarboxylase that intercepts MG-ACP rather than an isomerase acting on an *endo*-β-methyl branch.

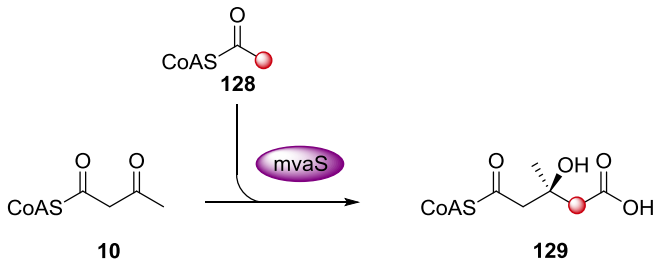
Finally, careful sequence analysis revealed a PGMG/PGAG motif that was conserved across ECH_2 from *trans*-AT PKS that either produce an *endo*- or an *exo*-β-methyl branch respectively. Homology modelling located these residues in a loop adjacent to the active site. In the *endo*-β-methyl producing ECH_2 domains, Met141 forms a narrow active site channel and His231 is hypothesised to both deprotonate prior to decarboxylation, and then reprotonate at the C-4 position to form the *endo*-β-methyl. In mECH, Ala138 allows for a more open active site in which His228 deprotonates and Tyr129 seems ideally positioned, proximal to the OAH-forming residues, to reprotonate at the C-2 position in the formation of an *exo*-β-methyl.

2.4.2 Future Work

There is considerable scope to further develop the work presented in this thesis. Structural characterisation of the ECH_2 domains (BatE and mECH) by X-ray crystallography would allow for validation of the active site configuration. This would provide valuable information into the key residues for catalysis and offer insight into the mechanism of selectivity. This could be further probed by rational engineering of the ECH_2 domains by point-mutation, crystallography and functional studies to understand the significance of the PGxG motif.

The NMR assay developed proved to be key in identifying and characterising the unsaturated β-methyl branches. The advantage of the assay was a single [^{13}C] label signal, formation of the correct stereoisomer of [^{13}C]-HMG-ACP and distinct chemical shifts and phase inversion by DEPT NMR at each stage of the β-branching pathway. The disadvantages of the assay were the time required in ACP loading and preparation, reliance on acetyl transfer in each reaction, substrate dilution by mixing of two ACPs and non-productive hydrolysis of [^{13}C]-Ac-BatC. These factors combined often led to a low flux of intermediates through the β-branching pathway.

A solution to this problem would be the pre-forming of [^{13}C]-HMG-CoA **129** for transfer onto the desired ACP. This could be achieved by reacting [^{13}C]-Ac-CoA **128** and Acac-CoA **10** in the presence of mvaS, the HMGCS domain from the mevalonate-dependant isoprenoid pathway (Scheme 36).[42,63] This reaction would produce the correct stereoisomer and might allow for the purification of [^{13}C]-HMG-CoA **129**. It may then be expediently loaded onto the desired ACP, giving a high proportion of the correct substrate with no dilution of the [^{13}C] label. This higher effective concentration gives greater sensitivity and would allow for shorter NMR experiments and better observation of short-lived species.



Scheme 36 Generation of [^{13}C]-HMG-CoA **129** would allow for direct loading of ACPs with an advanced biosynthetic intermediate, thus simplifying β-branching NMR assays.

3 Mupirocin/Thiomarinol Starter Unit Generation

3.1 Introduction to mupirocin/thiomarinol

Mupirocin is a mixture of pseudomonic acids A-C (PA-A to PA-C) that were first isolated from *Pseudomonas fluorescens* NCIMB 10586 in 1971,[147] with PA-A accounting for approximately 90% of the mixture (Figure 70A). The chemical structures of each component were reported in 1974, 1977 and 1980 respectively.[148–150] The carbon skeleton of each compound is related, with a polyketide-derived monic acid portion that is esterified onto a fatty-acid derived 9-hydroxynonanoic acid.[151]

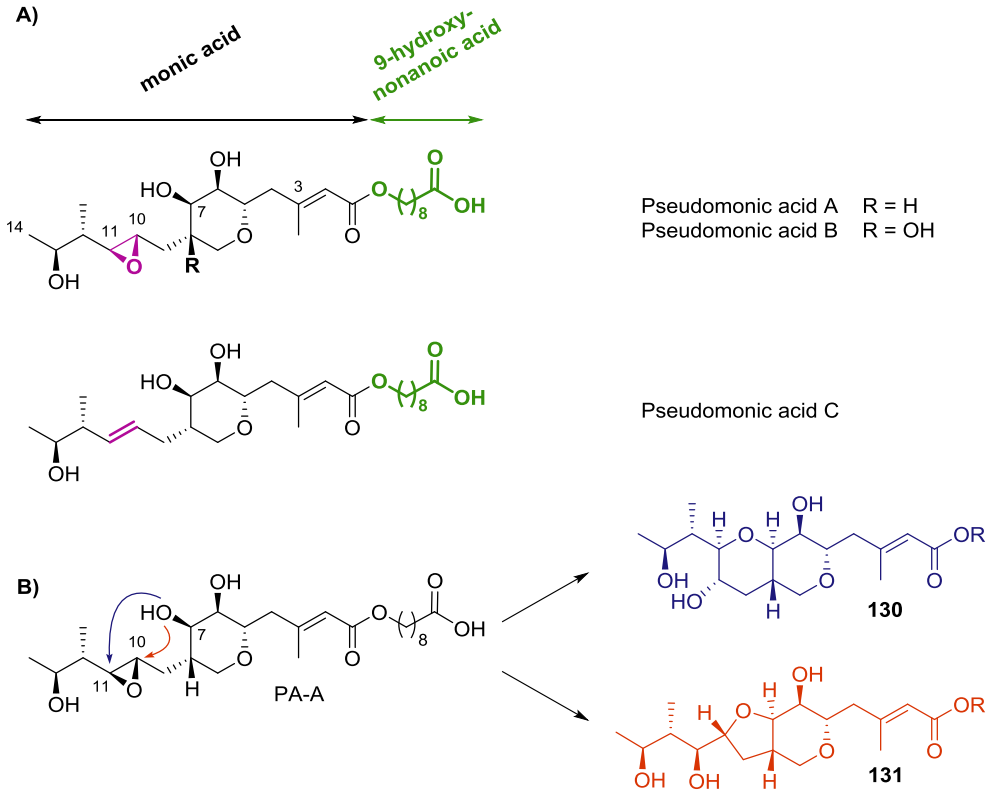


Figure 70 Mupirocin is a mixutre of pseudomonic acids A-C. A) The structures are composed of a PKS-derived monic acid portion esterified to a fatty acid-derived 9-hydroxynonanoic acid. B) Outside a narrow pH range (pH 4-9) the C-7 hydroxyl may attack the 10,11-epoxide forming inactive cyclic ethers **130** and **131**.

Mupirocin is a commercially available antibiotic used for the treatment of topical skin infection and marketed as Bactroban™. It is active against Gram positive bacteria and most significantly is used clinically to treat methicillin resistant *Staphylococcus aureus* (MRSA). Its wider application is limited by its poor bioavailability, a result of the metabolic instability of the compound when exposed to acidic conditions, at which point the C-7 hydroxyl group attacks the 10,11-epoxide forming antibiotic-inactive ethers **130** and **131** (Figure 70B).[152]

Analysis of the bioactivity of each component of mupirocin reveals a similar potency of PA-A and PA-C.[153] PA-C lacks the epoxide moiety of PA-A so does not suffer from this mode of acidic inactivation *in vivo*, however, it is only a minor component of the mixture. Increasing the titres of PA-C would allow for further study and possibly widening the application of this compound.

The antibacterial target for mupirocin is the bacterial isoleucyl-tRNA synthetase (IleRS), therefore inhibiting protein and RNA biosynthesis.[154] Co-crystallisation revealed that the C-14 region of PA-A mimicked isoleucine binding, whilst the THP ring and α,β-unsaturated ester occupy the ATP-binding pocket.[155,156] The fatty acid chain is important for complex stability and binds in a hydrophobic pocket.

A number of feeding experiments using different labelled isotopomers of acetate resulted in the incorporation of ^{18}O, ^{13}C and ^{2}H (Figure 71).[36,151] The labelling pattern was consistent with the head-to-tail condensation of acetate in the polyketide portion to form a heptaketide with SAM-mediated α-methylation at C-16 and C-17 and an acetate-derived *endo*-β-methyl branch at C-15. The fatty acid portion is proposed to be derived from a 3-carbon unit that has undergone three head-to-tail acetate condensations to give the C_9 acid. These labelling studies were one of the earliest reports demonstrating the origin of the C-15 β-branch to be the methyl group of a cleaved acetate.[36]

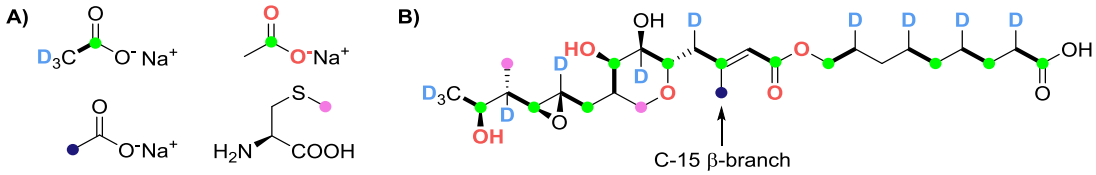


Figure 71 Isotopic feeding studies using the mupirocin WT producing strain *Pseudomonas fluorescens* NCIMB 10586. A) The acetate feedstocks with differential isotopic enrichment. B) The resulting labelling pattern of PA-A.

Mupirocin is assembled by a type I *trans*-AT PKS/FAS encoded by a 74 kb gene cluster consisting of 6 large ORFs, labelled mupirocin multifunctional proteins (MmpA to MmpF) and 29 smaller ORFs (MupA-X and MacpA-E) encoding for a range of tailoring enzymes.[157] MmpD and MmpA are the modular PKS components that synthesise the majority of the linear polyketide backbone (Figure 72). An HCS cassette interacts with the final di-domain ACP of MmpA to give an endo-β-methyl branched monic acid.

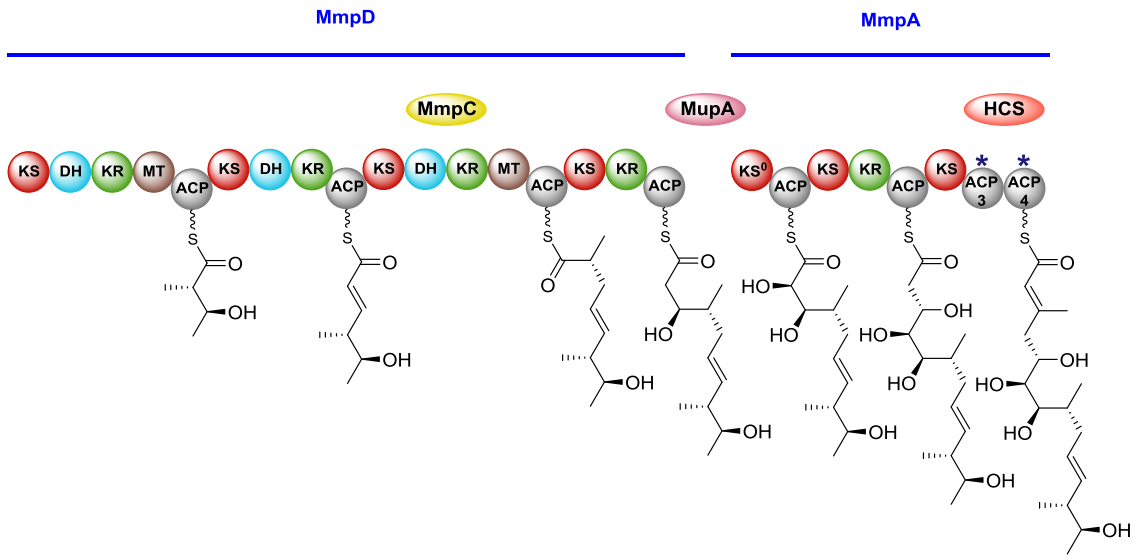
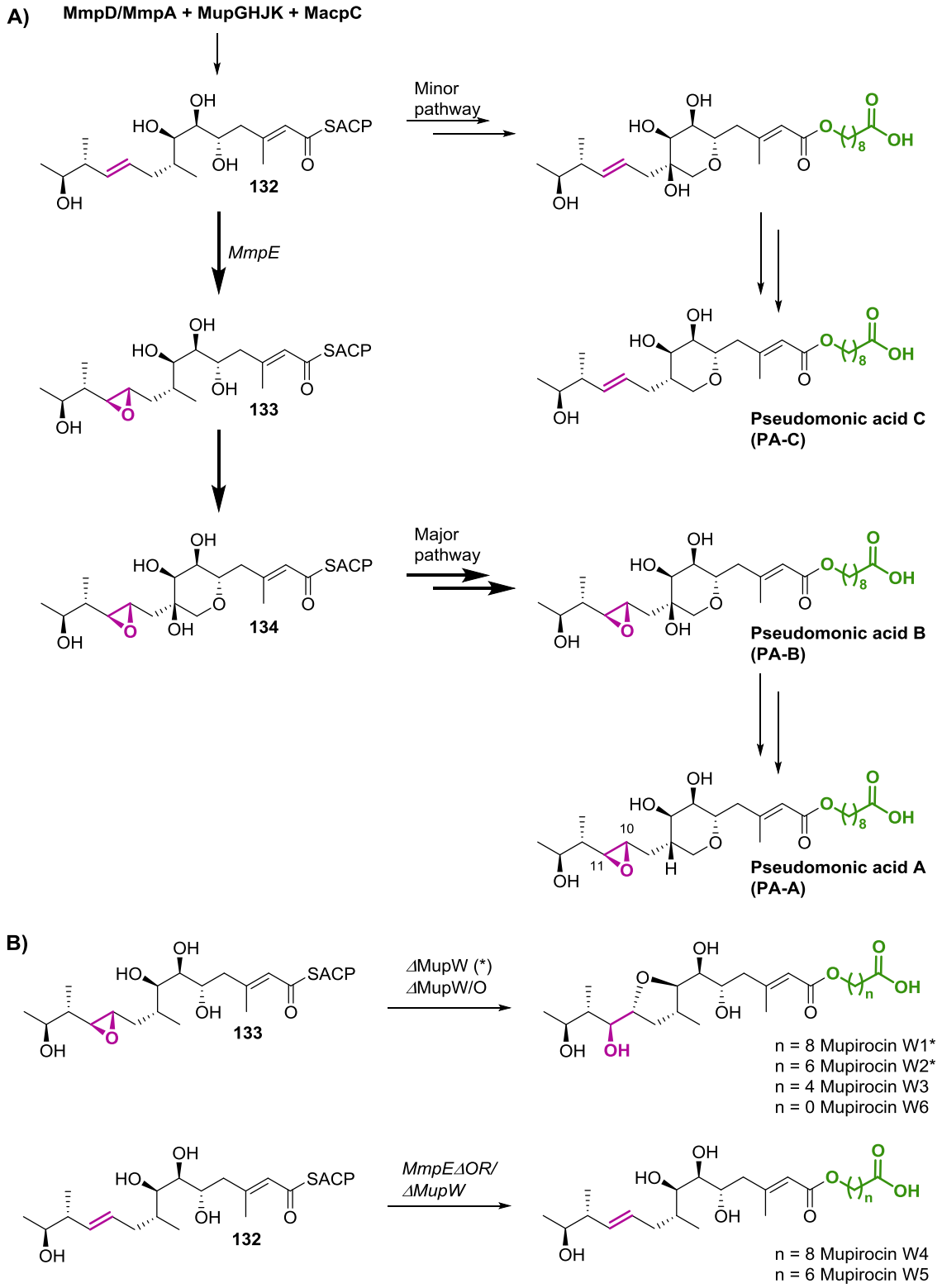


Figure 72 The modular assembly of the carbon skeleton of monic acid which is further modified by tailoring enzymes. *Trans*-acting enzymes provide additional functionality, including MmpC (ER), MupA (oxygenase) and a 5-enzyme HCS cassette (MacpA, MupGHJK) to install the *endo*-β-methyl branch.

The metabolic flux then diverges from **132** along parallel pathways (Scheme 37A). In the major pathway, epoxidation of the 10,11-alkene is catalysed by the C-terminal oxidoreductase of the di-domain MmpE.[158] THP ring formation is controlled by MupW and MupZ,[159] whilst esterification of the polyketide with 3-hydroxypropionate (3-HP) and three rounds of fatty acid biosynthesis furnishes PA-B.[160] The exact timing of esterification and fatty acid chain extension is not known and **133** and **134** are putative intermediates. Conversion of PA-B to the major metabolite PA-A requires a series of oxidative and reductive steps on the THP ring culminating in the removal of the C-8 hydroxyl group.[161] The minor pathway lacks the epoxidation step, however, THP formation, esterification and fatty acid biosynthesis, analogous to the major pathway, forms the minor component PA-C.[158]

The mupirocin biosynthetic pathway contains 11 type I ACPs that are part of the modular PKS as well as 5 type II ACPs. These free-standing ACPs (MacpA-E) are encoded by ORFs that lie in the tailoring region of the gene cluster and are essential for antibiotic production.[162] Each mutant strain (ΔMacpA-MacpE) could be complemented in-*trans* by its cognate ACP, however, no complementation was observed upon the overexpression of any noncognate ACPs.[118] Therefore each ACP has a defined role in the production of mupirocin.

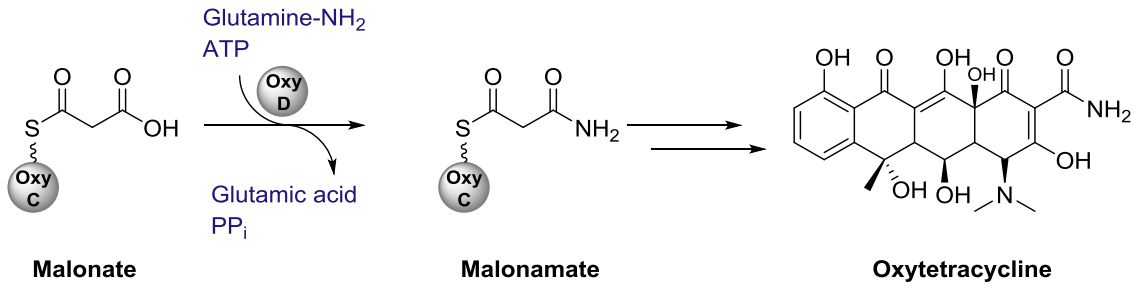


Scheme 37 A) Tailoring enzymes produce PA-A, *via* PA-B, along the major metabolic pathway, whilst PA-C is produced by a parallel, minor pathway. The ACP-bound species **133** and **134** are putative intermediates in which the alkene epoxidation (purple) is hypothesised to occur prior to esterification and fatty acid production (green). B) Mupirocins W1-W2 have been isolated from *ΔMupW* and mupirocins W3 and W6 from *ΔMupW/O*. Mupirocins W4-W5 have only been isolated from *MmpEΔOR/ΔMupW*.[158]

MacpC is the acetyl-donor ACP of the HCS cassette that introduces the β-methyl branch at C-15,[163] whilst MacpE is bound to PA-B during several of the steps in the conversion to PA-A.[164] The closest homologue of MacpD is DifC from the difficidin gene cluster and both ACPs lie adjacent to an acyl-CoA synthase (MupQ/DifD) and a 3-oxo-acyl ACP reductase (MupS/DifE) in their respective gene clusters. These cassettes are thought to be responsible for the formation of a 3-HP starter unit in mupirocin and an acrylate moiety in difficidin.[165,166]

The closest homologues of MacpA were ACPs linked to fatty acid biosynthesis whilst ACPs related to type I polyunsaturated fatty acid biosynthesis were the closest to MacpB.[118] It is therefore likely that one or both are involved in tethering the intermediates during the 9-hydroxynonanoic acid biosynthesis. An interesting feature of both MacpA and MacpD is the presence of an unusual C-terminal extension that may form a fifth helix. Shields and co-workers constructed MacpA/D mutant strains whereby 9 and 21 residues from the C-terminus of the ACP were deleted.[118] Mutated MacpA was unable to complement the WT ACP in a *ΔMacpA* strain. In contrast, mutated MacpD could complement the *ΔMacpD* strain, but with lower efficiency than the full-length ACP. The authors concluded that decreased protein affinity or disrupted protein structure result in the decrease in antibiotic production, demonstrating the importance of this final putative helix and suggesting it might be an important feature for enzyme recognition.

In the biosynthesis of oxytetracycline, an unusual malonamate starter unit is used to prime the PKS (Scheme 38). Tang *et al.* showed that OxyD, which shows homology to an asparagine synthetase, was responsible for the formation of malonamate.[167] The proposed mechanism is the activation of malonyl-OxyC (an ACP) by OxyD and ATP, followed by condensation with glutamine as the source of nitrogen. In 2003, the NMR structure for OxyC was solved by Crump *et al.* and showed, in addition to the typical 4 α-helices, a 13-residue C-terminal extension.[168] The similarity in proposed mechanism for malonamate in oxytetracycline and 3-HP in mupirocin biosynthesis suggests that the C-terminal extension of OxyC and MacpD is important for the interaction with OxyD and MupQ respectively.



Scheme 38 Malonamate starter unit generation in the biosynthesis of oxytetracycline. The ACP (OxyC) contains an unusual C-terminal extension similar to that observed for MacpA and MacpD.

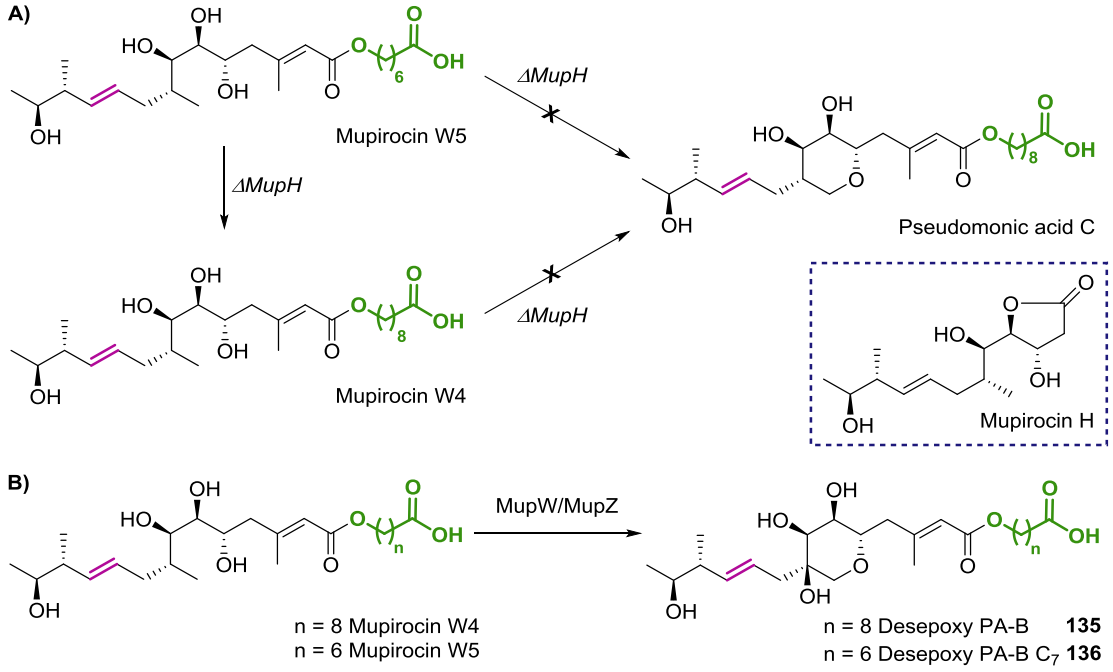
Timing of THP formation, esterification and fatty acid biosynthesis

Whilst some insights have been gained into the formation of the THP ring and the seemingly illogical conversion of PA-B to PA-A, the timing of the ring formation relative to fatty acid esterification and chain extension is still unknown. Conflicting evidence for the order of THP ring formation and esterification exist.

A single knock-out strain *ΔMupW*, and a double knock-out strain *ΔMupW/O* were generated and these yielded a range of isolable ring-opened compounds termed mupirocins W1-W6 (Scheme 37B).[158] Mupirocins W1-W3 and W6 are shunt products that arise from attack of the C-7 hydroxyl onto the 10,11-epoxide to generate a cyclic ether. W1-W3 contain C_9, C_7 and C_5 fatty acid chains respectively, whilst W6 lacks a fatty acid chain. Mupirocins W4 and W5 are the alkene equivalent (non-epoxidised) compounds with C_9 and C_7 fatty acid chains. These compounds demonstrated that esterification and chain extension of the ring-opened compounds was possible. However, this order of steps appears to be inefficient due to the formation of metabolites with varied chain lengths, suggesting that THP ring formation may happen prior to esterification.

Mupirocins W4 and W5 were fed into a strain that lacked the HMGS for β-methyl incorporation (*ΔMupH*) as shown in Scheme 39A.[158] The well documented phenotype of this strain is mupirocin H, arising from blocked flux along the type I pathway and cyclic ether formation.[169] It was hoped that W4 and W5 would be converted into PA-C, however, mupirocin W4 was not turned over and mupirocin W5 was only converted into W4 resulting from one round of fatty acid chain extension (Scheme 39A). These results support the hypothesis that once esterification and chain extension has taken place, epoxidation and THP ring formation is not permissive.

However, Wang and co-workers reported the reaction cascade for THP formation in mupirocin biosynthesis is catalysed by both MupW and MupZ.[159] Using a whole cell biotransformation, they demonstrated *in vitro* that MupW4 and W5 could be converted to desepoxy PA-B **135** and desepoxy PA-B C_7 **136** with a C_9 and C_7 fatty acid chain respectively (Scheme 39B). This result shows that THP ring formation may occur after esterification, creating uncertainty over the order of events in the WT pathway.



Scheme 39 Conflicting evidence for the timing of THP formation relative to esterification. A) Metabolites isolated from *P. fluorescens* knock-out strains suggest ring closure occurs prior to fatty acid chain extension. Feeding of mupirocin W4 and W5 to *ΔMupH* did not result in ring closure to form PA-C, but mupirocin W5 with a C_6 fatty acid chain was converted into W4 with a C_8 fatty acid chain. Structure of mupirocin H is shown in the blue box. B) mupirocin W4 and W5 were converted to the ring-closed products by incubating with MupW/MupZ in a whole cell biotransformation suggesting esterification may occur first.

Investigations into the function of MupR in the gene cluster determined its role as a regulator of the pathway.[170] When MupR was supplied in-*trans* to the *ΔMupR* strain, the whole pathway was upregulated, allowing for isolation of minor metabolites previously not observed during fermentation of the WT strain. Two compounds identified were **137** and **138** which are the C_7 derivatives of PA-A and PA-B respectively (Figure 73). The isolation of these products, along with mupirocins W2, W3 and W5 with a shortened fatty acid chain, supports the hypothesis that esterification occurs prior to chain extension. It is unlikely that esterification of shorter chain fatty acids occurs, rather, premature offloading from the fatty acid machinery likely forms the truncated compounds.

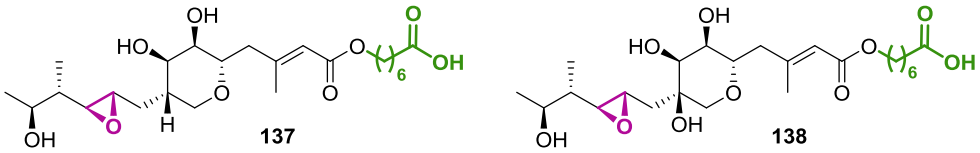


Figure 73 Shunt metabolites isolated from the in-*trans* over-expression of MupR in the wild-type *P. fluorescens* strain. Their isolation suggests the step-wise construction of the fatty acid portion following esterification.

Thiomarinol

Thiomarinol A is the major compound isolated from the marine bacterium *Pseudomonas* sp. SANK 73390 (Figure 74). Minor metabolites with varying oxygenation, alkylation and chain fatty acid chain length were also isolated from the WT organism.[171] They show promising antibiotic activity against Gram positive and Gram negative bacteria, including MRSA. The chemical structure consists of a polyketide-derived monic acid esterified onto 8-hydroxyoctanoic acid to form marinolic acid, a compound structurally related to mupirocin. Interestingly, thiomarinol contains the 10,11-alkene thus eliminating the *in vivo* metabolic instability observed for PA-A.[165] A pyrrothine moiety is then joined *via* an amide bond to furnish the thiomarinols, which are unusual, hybrid antibiotics with both the marinolic acid and pyrrothine portions displaying antibiotic activity.[172,173]

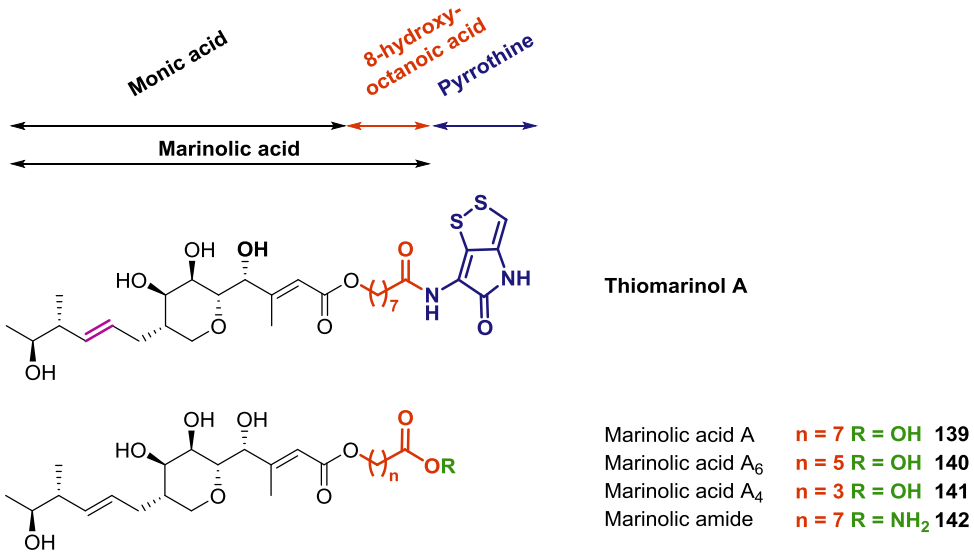


Figure 74 Structure of thiomarinol A isolated from *Pseudomonas* sp. SANK 73390 and marinolic acid/amide metabolites isolated from the WT and ΔNRPS strains.[171]

Feeding labelled acetates to the producing strain resulted in the expected head-to-tail condensation of acetate in the monic acid portion of thiomarinol A (Figure 75). The

oxygen atoms at C-1, C-5, C-7 and C-13 arise from acetate whereas those at C-4 and C-6 are derived from an alternative oxygen source. α-methylation by SAM is consistent with modular C-MeT domains, the C-15 β-methyl branch is observed from the cleaving of acetate by an HCS cassette and the pyrrothine moiety is constructed from two cysteine residues. However, an unexpected labelling pattern was observed for the C_8 fatty acid portion. A non-standard pattern and a reduced level of labelled acetate incorporation was observed for C5'-C-8' compared to the remainder of the molecule. Normal levels of incorporation at C-1' to C-4' were observed, suggesting a 4-carbon starter unit forms C-5' to C-8' that is subsequently extended by condensation of two head-to-tail acetate units. Succinyl CoA is a source of 4-hydroxybutyrate,[174] and by feeding [2,3-$^{13}C_2$]-succinate or [2,3-$^{13}C_2$]-4-hydroxybutyrate, an intact ^{13}C-^{13}C coupling constant was observed, identifying the source of the 4-carbon starter unit.

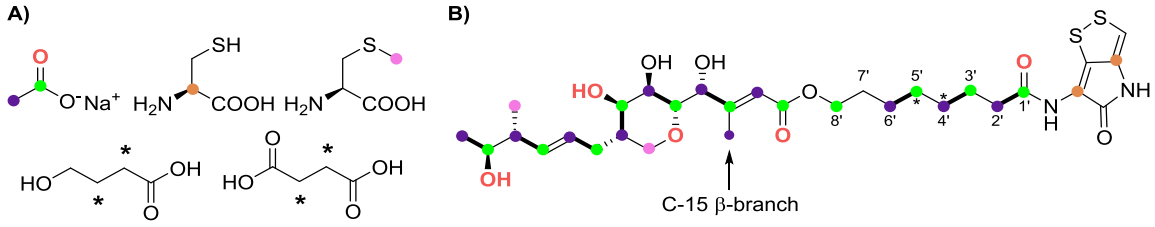
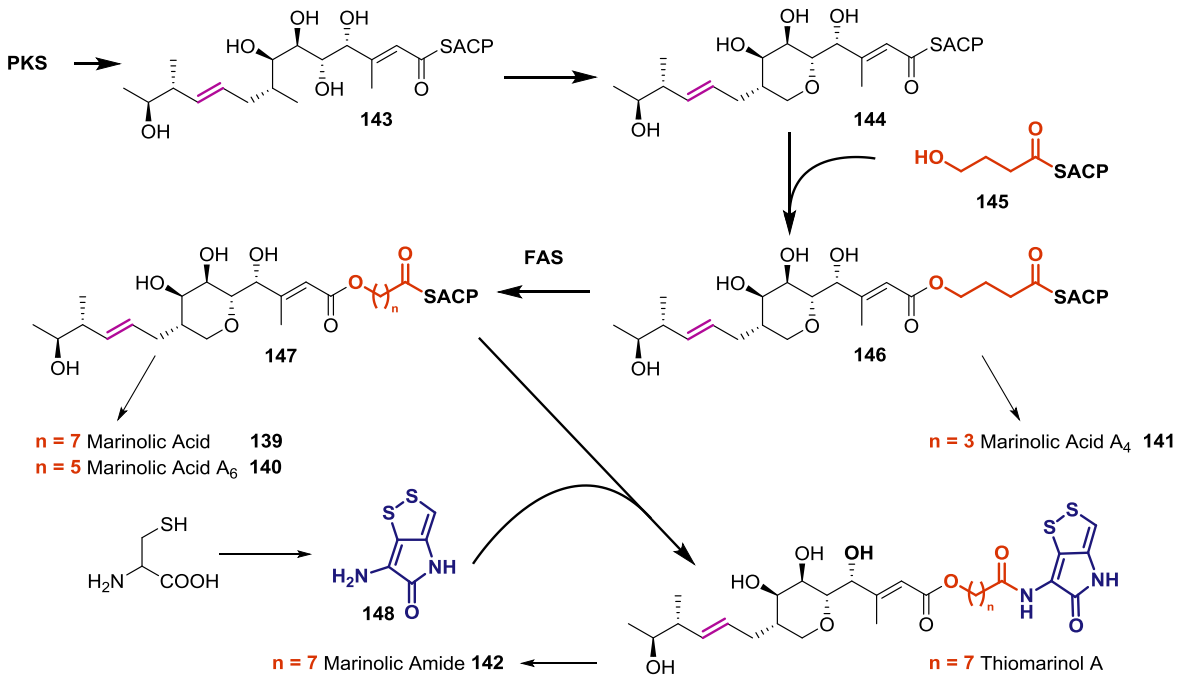


Figure 75 Isotopic feeding studies with the thiomarinol WT producing strain *Pseudomonas* sp. SANK 73390. A) Isotopically enriched precursors. B) The resulting labelling pattern of thiomarinol A where coloured dots correlate position of incorporation with the source of the [^{13}C] label. The asterisks (*) show the intact incorporation of 4-hydroxybutyrate, arising from succinate, into the fatty acid chain.

A 97 kb gene cluster encodes a hybrid PKS/FAS/NRPS with strong homology to the mupirocin gene cluster.[175] A modular PKS assembles the carbon backbone which is cleaved to give **143** and THP ring formation is then achieved in an analogous manner to mupirocin to give putative intermediate **144** (Scheme 40). ACP-bound 4-hydroxybutyrate (4-HB) **145** is most likely esterified prior to two rounds of fatty acid chain extension to give **147** and an amide condensation to install the pyrrothine moiety **148** give thiomarinol A. Minor metabolites have been isolated from the WT and *ΔNRPS* strains that feature marinolic acid with C_8 (**139**), C_6 (**140**) and C_4 (**141**) fatty acid chains as well as hydrolysis of the pyrrothine to give marinolic amide **142**. Reminiscent of the mupirocin pathway, the formation of these chain shortened compounds provides insight into the timing of esterification and extension, suggesting that ester-bond formation with the 4-carbon starter unit precedes FA biosynthesis.



Scheme 40 Biosynthesis of thiomarinol A, marinolic acids (**139**-**141**) and marinolic amide (**142**) minor metabolites.

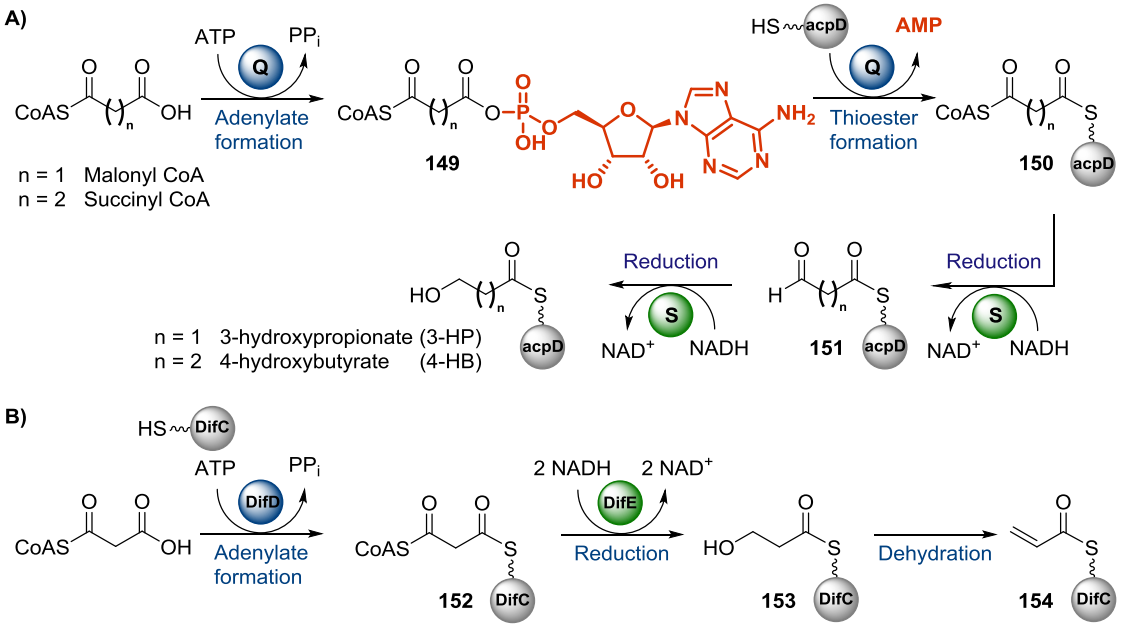
Starter unit generation

A conserved structural feature between mupirocin and thiomarinol A is the esterification of a polyketide-derived monic acid portion with a fatty acid. The starter unit in mupirocin is 3-hydroxypropionate which undergoes three rounds of fatty acid biosynthesis to give a C_9 chain, whereas thiomarinol starts with 4-hydroxybutyrate and undergoes two elongations to give a C_8 chain. These two fatty acid starter units are thought to be assembled by a three-enzyme cassette consisting of an ACP (M/TacpD), an acyl-CoA synthetase (MupQ/TmlQ) and a 3-oxo-reductase (MupS/TmlS).[165] Both sets of enzymes show significant sequence identity between the two pathways at 45.4%, 46.1% and 61.8% for the ACP, MupQ/TmlQ and MupS/TmlS respectively.

In mupirocin, it is hypothesised that MupQ selects malonyl-CoA and activates it as mixed anhydride **149**, prior to attack by *holo*-MacpD to form **150** (Scheme 41A). The domain type and mechanism are comparable to adenylation domains from NRPS that select amino acid extender units.[176,177] The unusual bis-thioester intermediate **150** is then reduced by MupS, which would deliver two equivalents of NADH selectively to the CoA thioester to give the 3-carbon alcohol *via* the aldehyde **151**. Thiomarinol proceeds *via* the same mechanistic pathway, except the 4-carbon starter unit is formed by the selection of succinyl CoA by TmlQ. This hypothesis was originally proposed through the observation of a homologous cassette of enzymes (DifC, DifD, DifE) in the difficidin biosynthetic

pathway that are proposed to form an acrylate starter unit **154** (Scheme 41B).[166] It is likely that the 3-carbon starter unit arises by the same mechanism as for mupirocin *via* **152**, however, the unsaturated moiety may be achieved by the dehydration of **153**.

The aim of this project was to investigate the mechanism of starter unit generation by *in vitro* reconstitution of the proposed pathway. This required the expression and purification of the appropriate *trans*-acting ACPs and catalytic enzymes, and reaction monitoring by MS, including the Ppant ejection assay.



Scheme 41 Starter unit generation by adenylation, thioesterification and subsequent reduction. A) MacpD, MupQ and MupS produce a 3-hydroxypropionate whilst TacpD, TmlQ and TmlS produce 4-hydroxybutyrate. B) A homologous cassette of enzymes in the biosynthesis of difficidin produces a 3-carbon acrylate starter unit.

3.2 Results and Discussion

3.2.1 Protein expression and purification

The genes encoding M/TacpA, TacpD, MupQ, MupS, TmlQ and TmlS were amplified from genomic DNA and cloned into POPINF whilst the gene for MacpD was synthesised and cloned into pET151. All proteins were transformed and expressed under standard conditions, yielding soluble protein that was purified by Ni^{2+}-affinity chromatography (Figures 76 and 77). ACPs were desalted into reaction buffer (50 mM Tris, 100 mM NaCl, pH 8.0) and stored at 4 °C or lyophilised and stored as white solids at -20 °C. MupQ/S and TmlQ/S were desalted into storage buffer, concentrated (~50 µM), flash frozen in aliquots and stored at -80 °C. Difficulty was experienced when concentrating MupQ/TmlQ due to loss of protein from precipitation and sticking to the membrane of

the spin concentrator. MupS/TmlS required the addition of 500 mM NaCl to the storage buffer and a minimum concentration of 10% glycerol to prevent complete precipitation.

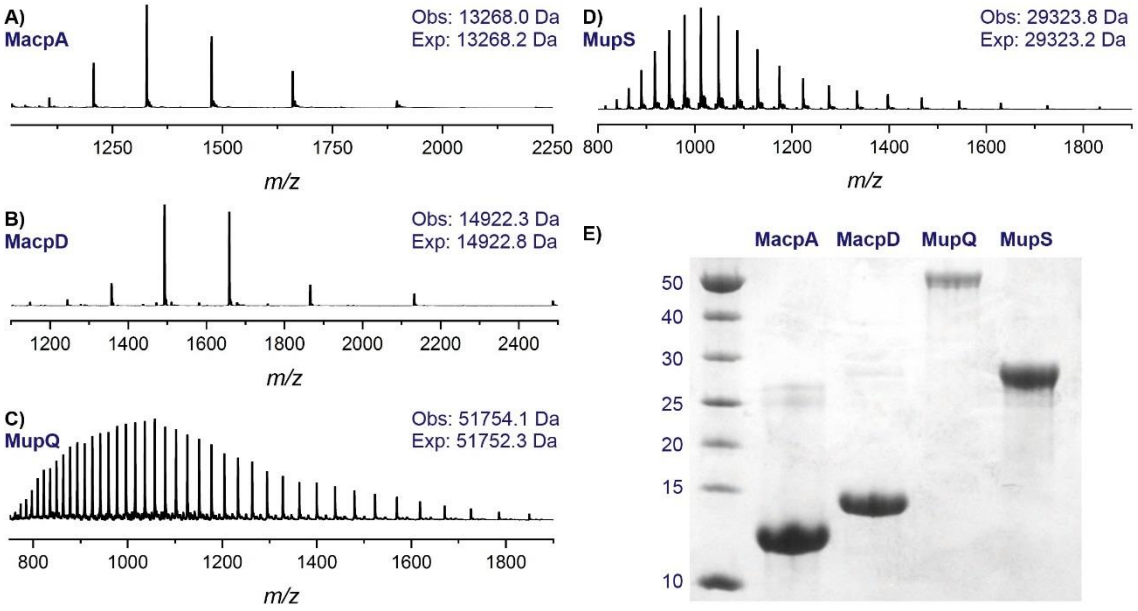


Figure 76 Characterisation of the protein cassette from the mupirocin gene cluster by SDS-PAGE and MS. All proteins contained an N-terminal His_6-tag. A) MS of apo-MacpA. B) MS of apo-MacpD. C) MS of MupQ. D) MS of MupS. E) SDS-PAGE analysis of apo-MacpA, apo-MacpD, MupQ and MupS confirmed proteins at the expected molecular mass.

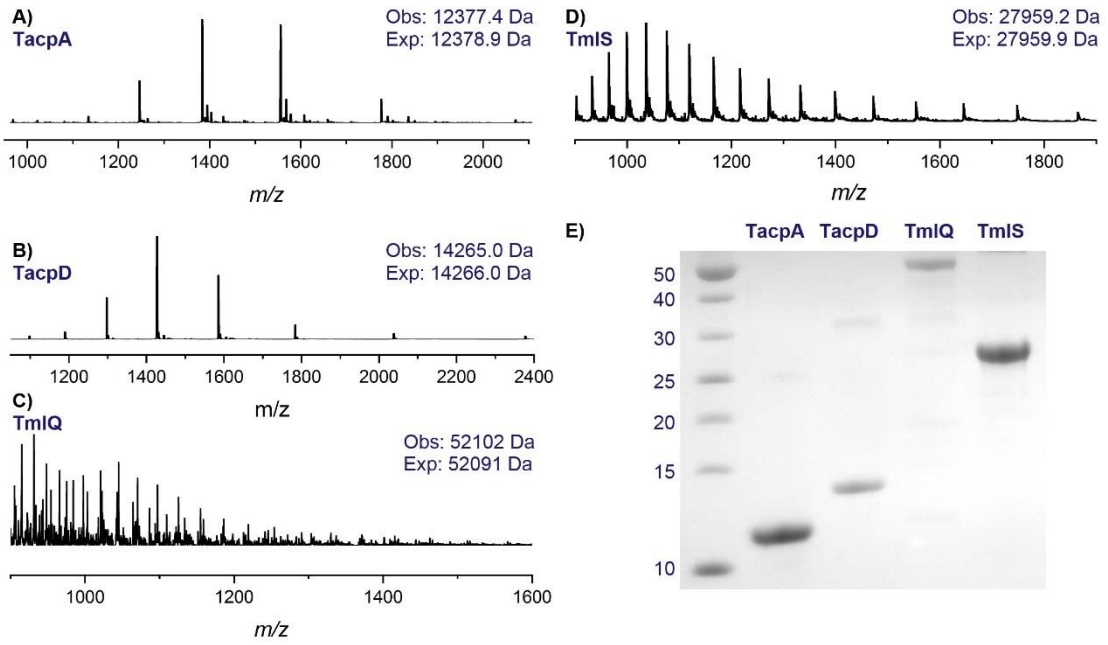
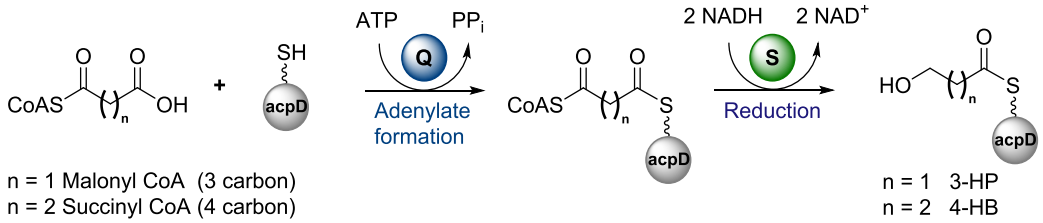


Figure 77 Characterisation of the protein cassette from the thiomarinol gene cluster by SDS-PAGE and MS. All proteins contained an N-terminal His_6-tag A) MS of apo-TacpA. B) MS of apo-TacpD. C) MS of TmlQ was difficult to analyse by MS and only a poor-quality spectrum with an 11 Da mass error could be obtained. D) MS of TmlS. E) SDS-PAGE analysis of apo-TacpA, apo-TacpD, TmlQ and TmlS confirmed proteins at the expected molecular mass.

M/TacpA and M/TacpD were all expressed as apo-ACP as confirmed by MS. Conversion to holo-ACP, by incubation with CoA, $MgCl_2$ and MupN, was monitored by MS and, upon completion, holo-ACP was desalted into reaction buffer supplemented with TCEP to remove excess CoA. The reaction could be carried out on a large scale (5-10 mg ACP) and the desalted protein was stored at -20 °C for up to 6 months.

3.2.2 *In vitro* starter unit generation

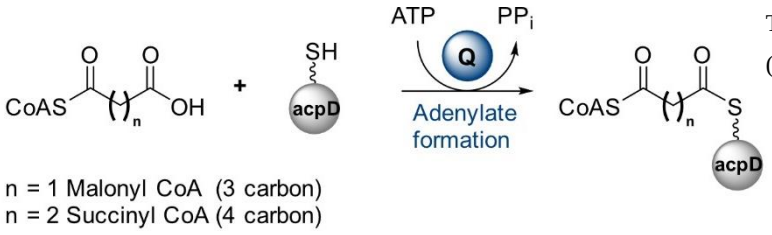


Scheme 42 Formation of a 3- or 4- carbon intermediate arising from the adenylation of malonyl or succinyl CoA respectively and subsequent attack of holo-ACP. The intermediate was then reduced by 2 equivalents of NADH to give 3-hydroxypropionate or 4-hydroxybutyrate, the proposed starter units in mupirocin and thiomarinol fatty acid biosynthesis respectively.

Having prepared the appropriate ACPs and catalytic domains, the reconstitution of the native mupirocin system to generate a 3-carbon starter unit was investigated (Scheme 42). Holo-MacpD was incubated with MupQ and malonyl CoA. MS analysis showed consumption of holo-MacpD and quantitative formation of a new species with a mass (16098 Da) consistent with the proposed 3-carbon intermediate attached *via* thioester linkages to both MacpD and CoA (Figure 78, entry 1 and B). The native thiomarinol system was then reconstituted *in vitro* by incubating holo-TacpD with succinyl CoA and TmlQ giving complete conversion to a 4-carbon intermediate with a mass of 15455 Da (Figure 78, entry 5 and F).

Having shown the formation of the 3- or 4- carbon intermediates bound between ACP and CoA thioesters, the selectivity of the adenylation step was investigated by varying the ACP, the adenylation domain and the CoA. MacpD was incubated with succinyl CoA and TmlQ which resulted in the quantitative formation of a 4-carbon activated intermediate with a mass of 16110 Da (Figure 78, entry 4 and C). Similarly, TacpD was incubated with malonyl CoA and MupQ which gave complete conversion to the respective 3-carbon intermediate (Figure 78, entry 8 and E). These results demonstrate the ability of both M/TacpD to operate with the non-native adenylation domain.

Entry	ACP	Co-factor	Q	Adenylation
1	MacpD	Mal	MupQ	✓ (B)
2		Succ	MupQ	×
3		Mal	TmlQ	×
4		Succ	TmlQ	✓ (C)
5	TacpD	Succ	TmlQ	✓ (F)
6		Mal	TmlQ	×
7		Succ	MupQ	×
8		Mal	MupQ	✓ (E)
9	MacpA	Mal	MupQ	×
10		Succ	TmlQ	×
11	TacpA	Succ	TmlQ	×
12		Mal	MupQ	×
13	BatA	Mal	MupQ	×



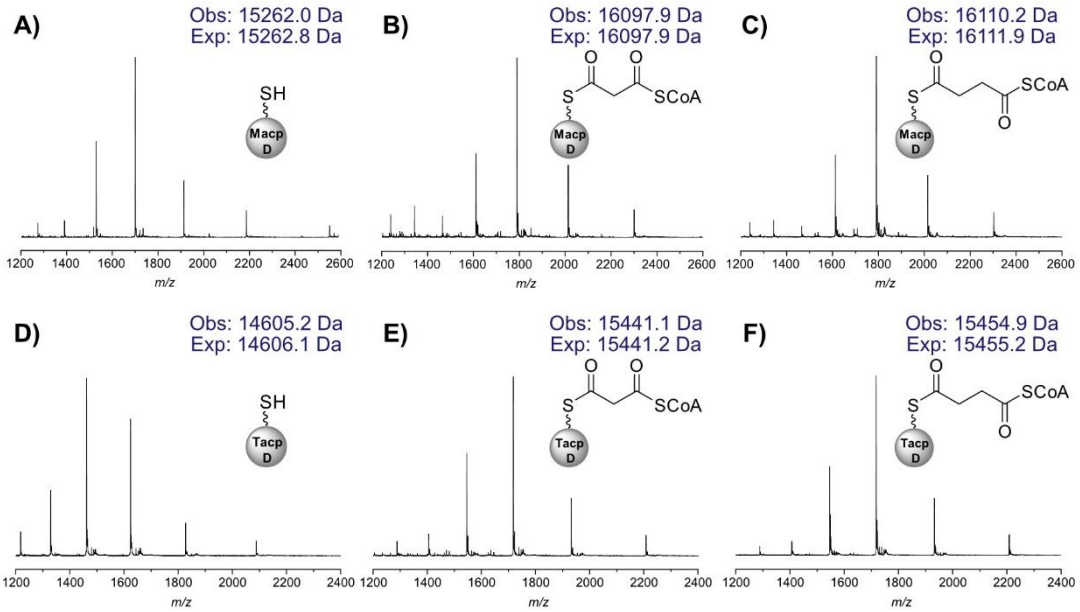
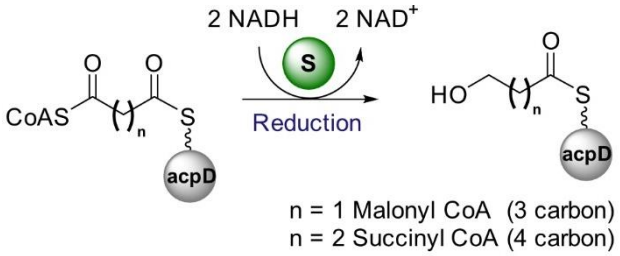


Figure 78 Reaction of M/TacpD with native and non-native adenylation domains. Successful reactions occurred for the correct co-factor/adenylation domain pair (MupQ + Mal-CoA or TmlQ + Succ-CoA) to introduce a 3- or 4-carbon unit respectively. M/TAcpA and BatA were unable to complement for M/TacpD with either pair. A) Holo-MacpD. B) Holo-MacpD + MupQ + Mal-CoA (entry 1). C) Holo-MacpD + TmlQ + Succ-CoA (entry 4). D) Holo-TacpD. E) Holo-TacpD + MupQ + Mal-CoA (entry 8). F) Holo-TacpD + TmlQ + Succ-CoA (Entry 5).

Incubating MacpD with either succinyl CoA/MupQ (Figure 78, entry 2) or malonyl CoA/TmlQ (Figure 78, entry 3) resulted in no intermediate formation, however, non-productive trans-thioesterification of the malonyl/succinyl group from CoA to holo-MacpD was observed. The same mis-match of adenylation domain and co-factor was investigated for TacpD, but once again trans-thioesterification was the only product

(Figure 78, entry 6 and 7). These results show that the adenylation domain is responsible for substrate selection and no promiscuity is observed. In a highly specific manner, malonyl CoA is selected and activated by MupQ, whilst succinyl CoA is selected by TmlQ. These findings correlate well with the selectivity of adenylation domains from NRPS pathways which are responsible for activating the correct amino acid as the extender unit in the construction of the peptides.[178,179]

Entry	ACP	Adenylation	S	Reduction
1	MacpD	3 carbon	MupS	✓ (A)
2			TmlS	×
3		4 carbon	TmlS	×
4			MupS	×
5	TacpD	4 carbon	TmlS	✓ (B)
6			MupS	×
7		3 carbon	MupS	✓ (C)
8			TmlS	✓ (C)



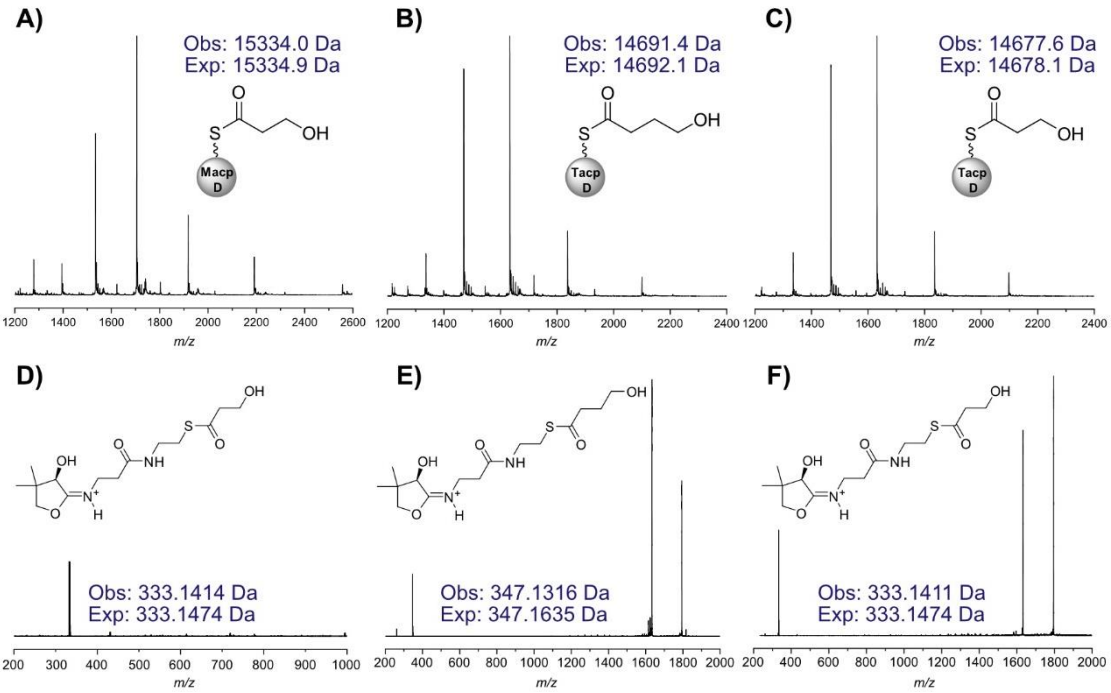


Figure 79 Reduction of 3- or 4-carbon adenylate intermediates by native or non-native reductase domains. A) MacpD + MupQ + Mal-CoA + MupS + NADH (entry1). B) TacpD + TmlQ + Succ-CoA + TmlS + NADH (Entry 5). C) TacpD + MupQ + Mal-CoA + MupS/TmlsS + NADH (entries 7 or 8). D) Ppant ejection for A. E) Ppant ejection for B. F) Ppant ejection for C.

Having shown the formation of a 3- and 4-carbon adenylate intermediate on both MacpD and TacpD was achieved by the correct pairing of adenylation domain and CoA, the subsequent reduction by MupS/TmlS was investigated. Addition of MupS and NADH to the 3-carbon intermediate bound to MacpD resulted in the reduction to a species consistent with the formation of MacpD-3-HP (Figure 79, entry 1 and A). To confirm reduction to the alcohol rather than the aldehyde, the product was confirmed by Ppant ejection that gave a peak at 333.14 Da (Figure 79D). TacpD-4-HB was generated by adding TmlS and NADH to the 4-carbon intermediate bound to TacpD (Figure 79, entry 5 and B), which was confirmed by Ppant ejection assay (Figure 79E).

Next, the promiscuity of the reductase domains was investigated. Entries 4 and 6 (Figure 79) show that MupS cannot reduce the 4-carbon unit regardless of whether it is attached to MacpD or TacpD. However, MupS was able to reduce the 3-carbon intermediate attached to TacpD (Figure 79, entry 7). TmlS was able to reduce the 3-carbon intermediate attached to the cognate TacpD (Figure 79, entry 8, C and F) but failed to reduce either the 3- or 4- carbon intermediate attached to MacpD (Figure 79, entry 2 and 3).

Finally, to confirm the previously reported inability of MacpA to complement MacpD *in vivo*,[118] MacpA was incubated with malonyl CoA/MupQ or succinyl CoA/TmlQ under standard conditions (Figure 78, entry 9 and 10). No evidence of the desired bis-thioester intermediate was detected by MS and the only product observed was trans-thioesterification of the CoA-bound acyl group to the thiol of holo-MacpA (Figure 80A and C). Repeating the reactions with TacpA in place of MacpA yielded the same result, as once again only thioesterification onto holo-TacpA was observed (Figure 80B and D). Further indication of the tight ACP-adenylation domain specificity was confirmed by the inability of BatA, a 4-helix, free-standing ACP from the kalimantacin biosynthetic pathway, to react with malonyl CoA/MupQ (Figure 78, entry 13). These results show specific interaction of M/TacpD with the adenylation domains and either malonyl or succinyl CoA is vitally important for the correct fidelity through the pathway.

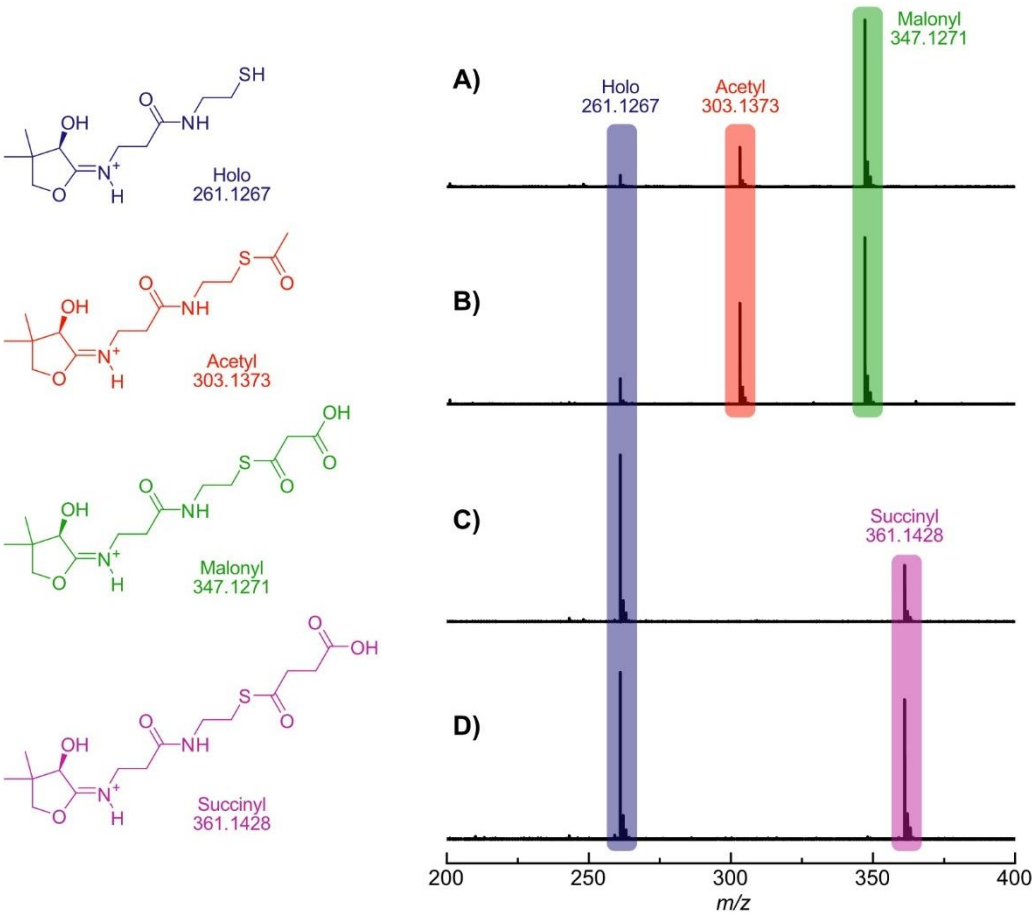
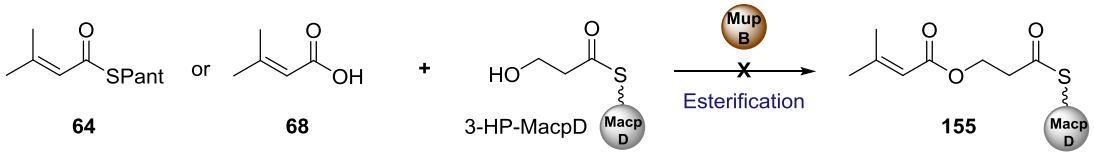


Figure 80 Ppant ejection ions shows trans-thioesterification rather than adenylation when M/TacpA was used rather than M/TacpD. When malonyl CoA is used, decarboxylation was also observed and results in the detection of acetyl-Ppant. A) MacpA + malonyl CoA + MupQ produces malonyl-MacpA (347.2 Da, green) and acetyl-MacpA (303.2 Da, red). B) TacpA + malonyl CoA + MupQ produced malonyl-TacpA (347.2 Da, green) and acetyl-TacpA (303.2 Da, red). C) MacpA + succinyl CoA + TmlQ produced succinyl-MacpA (361.2 Da, purple). D) TacpA + succinyl CoA + TmlQ produced succinyl-TacpA (361.2 Da, purple).

To test whether MupB was a candidate for the esterification of a short chain fatty acid onto the monic acid, it was cloned, expressed and purified. Freshly generated MacpD-3HP was incubated with enoyl-pantetheine **64** or enoyl-acid **68** and MupB (Scheme 43). The reaction was monitored by MS, however, no evidence of esterification product **155** was observed. Uncertainty exists over whether MupB would act as the ester-forming enzyme and due to time constraints further work was not undertaken on this reaction.

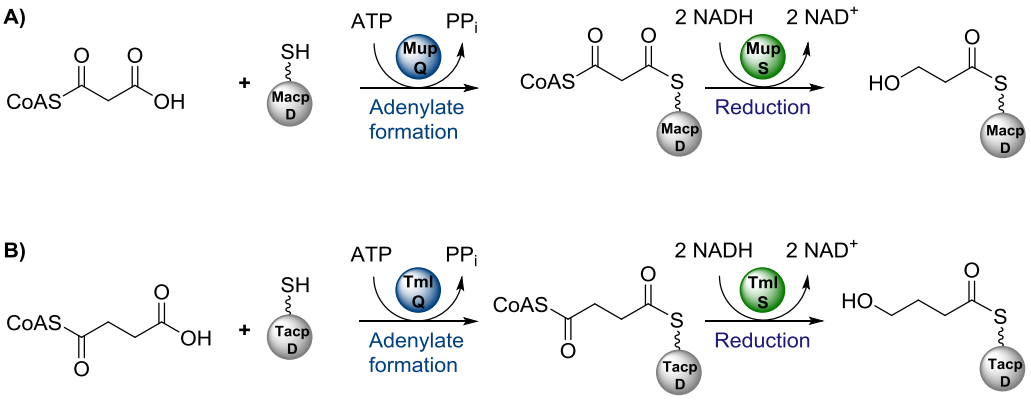


Scheme 43 MupB-catalysed esterification of MacpD-3HP was not possible with pantetheine **64** or acid **68**.

3.3 Conclusions and Future Work

3.3.1 Conclusions

The cassette of enzymes hypothesised to be responsible for the generation of a 3-HP and 4-HB starter unit in the biosynthesis of mupirocin and thiomarinol respectively consists of an ACP (M/TacpD), an adenylation domain (MupQ/TmlQ) and a reductase (MupS/TmlS). Each of these proteins was expressed, purified and characterised by SDS-PAGE and MS. The proposed pathway was then reconstituted *in vitro* and proved the formation of 3-HP and 4-HB (Scheme 44). In mupirocin biosynthesis, malonyl-CoA is activated by MupQ and ATP to form an adenylate intermediate, which is then trapped by holo-MacpD. The resulting three-carbon bis-thioester is subsequently reduced by MupS and 2 equivalents of NADH to form MacpD-3-HP. In thiomarinol biosynthesis, holo-TacpD receives activated succinyl-CoA from TmlQ, and the resulting four-carbon intermediate is reduced by TmlS and NADH to form TacpD-4-HB. The adenylation domains (MupQ/TmlQ) were shown to be selective for malonyl- and succinyl-CoA respectively, whereas the reductase domains (MupS/TmlS) showed some promiscuity for ACP identity and chain length.



Scheme 44 Mupirocin/thiomarinol starter unit generation. A) 3-HP formation by MacpD, MupQ and MupS from the mupirocin gene cluster. B) 4-HB formation by TacpD, TmlQ and TmlS from the thiomarinol gene cluster.

3.3.2 Future Work

Further work aims to further structurally characterise the enzymes of the pathway. X-ray crystallography structures of the adenylation (MupQ/TmlQ) and reductase (MupS/TmlS) domains would offer insight in to the selectivity observed. Particularly interesting would be the active site differences between MupQ and TmlQ that allows for the strict selection of CoA derivatives that differ by only CH_2. This could be further probed by mutagenesis of active site residues to potentially alter the domain selectivity.

The C-terminal extension of M/TacpD was highlighted as an unusual feature, but one that was also been observed in an ACP (oxyC)/adenylation (oxyD) domain pair in the biosynthesis of oxytetracyclines.[167,168] Understanding the secondary structure formed by these extensions and how they mediate protein-protein interactions would be important for future engineering of the pathway. NMR studies should be undertaken to understand whether this extension forms any secondary structure and whether this is influenced by the addition of MupQ/TmlQ or MupS/TmlS. A C-terminal extension was also observed in M/TacpA, however, these ACPs were unable to complement for M/TacpD. The role of the free-standing ACPs M/TacpA as well as M/TacpB (previously insoluble[118]) are currently unknown and may provide insight into substrate delivery and timing in the latter stages of mupirocin/thiomarinol biosynthesis.

Finally, the identity of the enzyme that catalyses esterification is not known, although candidates do exist in the gene cluster (MupB/MupL). The timing of post-PKS processing steps, including the order of esterification, fatty acid chain extension and THP ring formation is still uncertain.

4 References

1 J. O'Neill, *The Review On Antimicrobial Resistance. Tackling Drug-Resistant Infections Globally: Final Report And Recommendations*, 2016.

2 M. A. Fischbach and C. T. Walsh, *Science*, 2009, **325**, 1089–1093.

3 C. T. Walsh and T. A. Wencewicz, *J. Antibiot. (Tokyo).*, 2014, **67**, 7–22.

4 C. Nathan, *Nature*, 2004, **431**, 899–902.

5 L. L. Silver, *Clin. Microbiol. Rev.*, 2011, **24**, 71–109.

6 A. Okano, N. A. Isley and D. L. Boger, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E5052–E5061.

7 F. Alberti, K. Khairudin, E. R. Venegas, J. A. Davies, P. M. Hayes, C. L. Willis, A. M. Bailey and G. D. Foster, *Nat. Commun.*, 2017, **8**, 1831.

8 A. M. Bailey, F. Alberti, S. Kilaru, C. M. Collins, K. De Mattos-Shipley, A. J. Hartley, P. Hayes, A. Griffin, C. M. Lazarus, R. J. Cox, C. L. Willis, K. O'Dwyer, D. W. Spence and G. D. Foster, *Sci. Rep.*, 2016, **6**, 1–11.

9 E. Tacconelli, I. B. Autenrieth and A. Peschel, *Science*, 2017, **355**, 689–690.

10 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.

11 G. M. Cragg and D. J. Newman, *Biochim. Biophys. Acta - Gen. Subj.*, 2013, **1830**, 3670–3695.

12 N. Collie and W. S. Myers, *J. Chem. Soc., Trans.*, 1893, **63**, 122–128.

13 J. N. Collie, *J. Chem. Soc. Trans.*, 1907, **91**, 1806–1813.

14 A. J. Birch, R. A. Massy-Westropp and C. J. Moye, *Aust. J. Chem.*, 1955, **8**, 539–544.

15 R. E. Minto and C. A. Townsend, *Chem. Rev.*, 1997, **97**, 2537–2556.

16 J. Rinkel and J. S. Dickschat, *Beilstein J. Org. Chem.*, 2015, **11**, 2493–2508.

17 C. Hertweck, *Angew. Chemie Int. Ed.*, 2009, **48**, 4688–4716.

18 C. T. Walsh and M. A. Fischbach, *J. Am. Chem. Soc.*, 2010, **132**, 2469–2493.

19 E. J. N. Helfrich, S. Reiter and J. Piel, *Curr. Opin. Biotechnol.*, 2014, **29**, 107–115.

20 E. J. N. Helfrich and J. Piel, *Nat. Prod. Rep.*, 2016, **33**, 231–316.

21 A. T. Keatinge-Clay, *Chem. Rev.*, 2017, **117**, 5334–5366.

22 S. Smith and S.-C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.

23 B. Shen, *Curr. Opin. Chem. Biol.*, 2003, **7**, 285–295.

24 C. Hertweck, A. Luzhetskyy, Y. Rebets and A. Bechthold, *Nat. Prod. Rep.*, 2007, **24**, 162–190.

25 A. T. Keatinge-Clay, *Cell Chem. Biol.*, 2016, **23**, 540–542.

26 D. A. Herbst, C. A. Townsend and T. Maier, *Nat. Prod. Rep.*, 2018, **35**, 1046–1069.

27 J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.

28 L. Du, C. Sanchez and B. Shen, *Metab. Eng.*, 2001, **3**, 78–95.

29 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.

30 R. V. O'Brien, R. W. Davis, C. Khosla and M. E. Hillenmeyer, *J. Antibiot. (Tokyo).*, 2014, **67**, 89–97.

31 J. Staunton and B. Wilkinson, *Chem. Rev.*, 1997, **97**, 2611–2630.

32 C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr and D. E. Cane, *Annu. Rev. Biochem.*, 2007, **76**, 195–221.

33 S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–517.

34 J. R. Whicher, S. Dutta, D. A. Hansen, W. A. Hale, J. A. Chemler, A. M. Dosey, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 560–564.

35 J. Davison, J. Dorival, H. Rabeharindranto, H. Mazon, B. Chagot, A. Gruez and K.

J. Weissman, *Chem. Sci.*, 2014, **5**, 3081–3095.

36 F. M. Martin and T. J. Simpson, *J. Chem. Soc., Perkin Trans. 1*, 1989, 207–209.

37 D. G. I Kingston, M. X. Kolpak, J. W. LeFevre and I. Borup-Grochtmann, *J. Am. Chem. Soc.*, 1983, **105**, 5106–5110.

38 C. T. Calderone, *Nat. Prod. Rep.*, 2008, **25**, 845–853.

39 C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh and P. C. Dorrestein, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 8977–82.

40 L. Gu, J. Jia, H. Liu, K. Håkansson, W. H. Gerwick and D. H. Sherman, *J. Am. Chem. Soc.*, 2006, **128**, 9014–9015.

41 F. P. Maloney, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 10316–10321.

42 H. M. Miziorko, *Arch. Biochem. Biophys.*, 2011, **505**, 131–143.

43 C. T. Calderone, D. F. Iwig, P. C. Dorrestein, N. L. Kelleher and C. T. Walsh, *Chem. Biol.*, 2007, **14**, 835–846.

44 J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani and S. Matsunaga, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16222–16227.

45 L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V Grindberg, L. Gerwick, K. Håkansson, P. Wipf, J. L. Smith, W. H. Gerwick and D. H. Sherman, *Nature*, 2009, **459**, 731–735.

46 K. Bloch, *Science*, 1965, **150**, 19–28.

47 B. M. Lange, T. Rujan, W. Martin and R. Croteau, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 13172–13177.

48 B. W. Agranoff, H. Eggerer, U. Henning and F. Lynen, *J. Biol. Chem.*, 1960, **235**, 326–332.

49 T. J. Buchholz, C. M. Rath, N. B. Lopanik, N. P. Gardner, K. Håkansson and D. H. Sherman, *Chem. Biol.*, 2010, **17**, 1092–1100.

50 V. Simunovic and R. Müller, *ChemBioChem*, 2007, **8**, 497–500.

51 A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hothersall, E. Płoskoń, P. Wattana-amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi, J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas and M. P. Crump, *Nat. Chem. Biol.*, 2013, **9**, 685–692.

52 A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399–6408.

53 L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angew. Chemie Int. Ed.*, 2011, **50**, 2795–2798.

54 Z. Wang, S. R. Bagde, G. Zavala, T. Matsui, X. Chen and C. Y. Kim, *ACS Chem. Biol.*, 2018, **13**, 3072–3077.

55 T. W. Geders, L. Gu, J. C. Mowers, H. Liu, W. H. Gerwick, K. Håkansson, D. H. Sherman and J. L. Smith, *J. Biol. Chem.*, 2007, **282**, 35954–35963.

56 J. Piel, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14002–14007.

57 A. Kampa, A. N. Gagunashvili, T. A. M. Gulder, B. I. Morinaka, C. Daolio, M. Godejohann, V. P. W. Miao, J. Piel and O. S. Andresson, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E3129–E3137.

58 A. Nakabachi, R. Ueoka, K. Oshima, R. Teta, A. Mangoni, M. Gurgui, N. J. Oldham, G. Van Echten-Deckert, K. Okamura, K. Yamamoto, H. Inoue, M. Ohkuma, Y. Hongoh, S. Y. Miyagishima, M. Hattori, J. Piel and T. Fukatsu, *Curr. Biol.*, 2013, **23**, 1478–1484.

59 K. M. Fisch, C. Gurgui, N. Heycke, S. A. Van Der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews and J. Piel, *Nat. Chem. Biol.*, 2009, **5**, 494–501.

60 A. S. Eustáquio, J. E. Janso, A. S. Ratnayake, C. J. O'Donnell and F. E. Koehn, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, 1–10.

61 X. Liu, S. Biswas, M. G. Berg, C. M. Antapli, F. Xie, Q. Wang, M. C. Tang, G. L. Tang, L. Zhang, G. Dreyfuss and Y. Q. Cheng, *J. Nat. Prod.*, 2013, **76**, 685–693.

62 J. C. Kwan, M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood and E. W. Schmidt, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20655–20660.

63 S. T. Slocum, A. N. Lowell, A. N. Tripathi, V. V. Shende, J. L. Smith and D. H. Sherman, *Methods Enzymol.*, 2018, **604**, 207–236.

64 B. Rohm, K. Scherlach and C. Hertweck, *Org. Biomol. Chem.*, 2010, **8**, 1520–1522.

65 N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth and C. Hertweck, *Chem. Biol.*, 2012, **19**, 1164–1174.

66 D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 14234–14243.

67 L. Song, M. Jenner, J. Masschelein, C. Jones, M. J. Bull, S. R. Harris, R. C. Hartkoorn, A. Vocat, I. Romero-Canelon, P. Coupland, G. Webster, M. Dunn, R. Weiser, C. Paisey, S. T. Cole, J. Parkhill, E. Mahenthiralingam and G. L. Challis, *J. Am. Chem. Soc.*, 2017, **139**, 7974–7981.

68 M. J. Bertin, A. Vulpanovici, E. A. Monroe, A. Korobeynikov, D. H. Sherman, L. Gerwick and W. H. Gerwick, *ChemBioChem*, 2016, **17**, 164–173.

69 W. Trowitzsch-Kienast, K. Schober, V. Wray, K. . Gerth, H. Reichenbach and G. Hefle, *Liebigs Ann. Chem.*, 1989, 345–355.

70 V. Simunovic and R. Müller, *ChemBioChem*, 2007, **8**, 1273–1280.

71 R. J. Cox, *ChemBioChem*, 2014, **15**, 27–29.

72 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher and C. Hertweck, *Nature*, 2013, **502**, 124–128.

73 D. Heine, T. Bretschneider, S. Sundaram and C. Hertweck, *Angew. Chemie Int. Ed.*, 2014, **53**, 11645–11649.

74 S. Sundaram, D. Heine and C. Hertweck, *Nat. Chem. Biol.*, 2015, **11**, 949–951.

75 K. Kamigiri, Y. Suzuki, M. Shibazaki, M. Morioka, K. Suzuki, T. Tokunaga, B. Setiawan and R. M. Rantiatmodjo, *J. Antibiot. (Tokyo).*, 1996, **49**, 136–139.

76 T. Tokunaga, K. Kamigiri, M. Orita, T. Nishikawa, M. Shimizu and H. Kaniwa, *J. Antibiot. (Tokyo).*, 1996, **49**, 140–144.

77 V. V. Smirnov, L. N. Churkina, V. I. Perepnikhatka, N. S. Mukvich, A. D.

Garagulia, E. A. Kiprianova, A. N. Kravets and S. A. Dovzhenko, *Prikl. Biokhim. Mikrobiol.*, 2000, **36**, 55–58.

78 W. Mattheus, L. J. Gao, P. Herdewijn, B. Landuyt, J. Verhaegen, J. Masschelein, G. Volckaert and R. Lavigne, *Chem. Biol.*, 2010, **17**, 149–159.

79 V. V Klochko, E. A. Kiprianova, L. N. Churkina and L. V Avdeeva, *Mikrobiolohichnyĭ zhurnal (Kiev, Ukr. 1993)*, 2008, **70**, 41–46.

80 V. V. Klochko, L. B. Zelena, J. Y. Kim, L. V. Avdeeva and O. N. Reva, *Int. J. Antimicrob. Agents*, 2016, **47**, 56–61.

81 M. Jenner, J. Afonso, C. Kohlhaas, P. Karbaum, S. Frank, J. Piel and N. J. Oldham, *Chem. Commun.*, 2016, **52**, 1–4.

82 D. A. Vander Wood and A. T. Keatinge-Clay, *Proteins Struct. Funct. Bioinforma.*, 2018, **86**, 664–675.

83 L. Zhang, T. Hashimoto, B. Qin, J. Hashimoto, I. Kozone, T. Kawahara, M. Okada, T. Awakawa, T. Ito, Y. Asakawa, M. Ueki, S. Takahashi, H. Osada, T. Wakimoto, H. Ikeda, K. Shin-ya and I. Abe, *Angew. Chemie Int. Ed.*, 2017, **56**, 1740–1745.

84 M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chemie Int. Ed.*, 2013, **52**, 1143–1147.

85 R. J. Heath and C. O. Rock, *J. Biol. Chem.*, 1995, **270**, 26538–26542.

86 R. Heath, S. White and C. Rock, *Appl. Microbiol. Biotechnol.*, 2002, **58**, 695–703.

87 V. E. Lee and A. J. O'Neill, *Int. J. Antimicrob. Agents*, 2017, **49**, 121–122.

88 W. Mattheus, J. Masschelein, L. J. Gao, P. Herdewijn, B. Landuyt, G. Volckaert and R. Lavigne, *Chem. Biol.*, 2010, **17**, 1067–1071.

89 S. Brinster, G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss and C. Poyart, *Nature*, 2009, **458**, 83–86.

90 W. Balemans, N. Lounis, R. Gilissen, J. Guillemont, K. Simmen, K. Andries and A. Koul, *Nature*, 2010, **463**, E3–E4.

91 S. Brinster, G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss and C. Poyart, *Nature*, 2010, **463**, E4–E5.

92 E. Cundliffe, *Annu. Rev. Microbiol.*, 1989, **43**, 207–233.

93 H. Xu, L. Heide and S. M. Li, *Chem. Biol.*, 2004, **11**, 655–662.

94 U. Galm, L. Wang, E. Wendt-Pienkowski, R. Yang, W. Liu, M. Tao, J. M. Coughlin and B. Shen, *J. Biol. Chem.*, 2008, **283**, 28236–28245.

95 V. M. D'Costa, K. M. McGrann, D. W. Hughes and G. D. Wright, *Science*, 2006, **311**, 374–377.

96 I. R. G. Thistlethwaite, F. M. Bull, C. Cui, P. D. Walker, S. S. Gao, L. Wang, Z. Song, J. Masschelein, R. Lavigne, M. P. Crump, P. R. Race, T. J. Simpson and C. L. Willis, *Chem. Sci.*, 2017, **8**, 6196–6201.

97 I. Thistlethwaite, PhD Thesis, University of Bristol, 2015.

98 F. M. Bull, PhD Thesis, University of Bristol, 2016.

99 C. T. Lohans, D. Y. Wang, J. Wang, R. B. Hamed and C. J. Schofield, *ACS Catal.*, 2017, **7**, 6587–6599.

100 R. B. Hamed, E. T. Batchelar, I. J. Clifton and C. J. Schofield, *Cell. Mol. Life Sci.*, 2008, **65**, 2507–2527.

101 H. M. Holden, M. M. Benning, T. Haller and J. A. Gerlt, *Acc. Chem. Res.*, 2001, **34**, 145–157.

102 G. Agnihotri and H. W. Liu, *Bioorganic Med. Chem.*, 2003, **11**, 9–20.

103 L. Hedstrom, *Chem. Rev.*, 2002, **102**, 4501–4523.

104 H.-H. Otto and T. Schirmeister, *Chem. Rev.*, 1997, **97**, 133–172.

105 F. Zhang, H. Y. He, M. C. Tang, Y. M. Tang, Q. Zhou and G. L. Tang, *J. Am. Chem. Soc.*, 2011, **133**, 2452–2462.

106 H. Xiang, L. Luo, K. L. Taylor and D. Dunaway-Mariano, *Biochemistry*, 1999, **38**, 7638–7652.

107 T. R. Kiema, C. K. Engel, W. Schmitz, S. A. Filppula, R. K. Wierenga and J. K. Hiltunen, *Biochemistry*, 1999, **38**, 2991–2999.

108 W. Yu, X. Chu, G. Deng, X. Liu, G. Chen and D. Li, *Biochim. Biophys. Acta* -

Gen. Subj., 2006, **1760**, 1874–1883.

109 D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts and W. H. Gerwick, *Chem. Biol.*, 2004, **11**, 817–833.

110 J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, **29**, 1111–1137.

111 J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel and M. D. Burkart, *Nat. Prod. Rep.*, 2014, **31**, 61–108.

112 R. J. Cox, J. Crosby, O. Daltrop, F. Glod, M. E. Jarzabek, T. P. Nicholson, M. Reed, T. J. Simpson, L. H. Smith, F. Soulas, A. E. Szafranska and J. Westcott, *J. Chem. Soc. Perkin Trans. 1*, 2002, **14**, 1644–1649.

113 P. Dall Aglio, C. J. Arthur, C. Williams, K. Vasilakis, H. J. Maple, J. Crosby, M. P. Crump and A. T. Hadfield, *Biochemistry*, 2011, **50**, 5704–5717.

114 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923–936.

115 L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585–1595.

116 K. Reuter, M. R. Mofid, M. A. Marahiel and R. Ficner, *EMBO J.*, 1999, **18**, 6823–6831.

117 K. Finzel, D. J. Lee and M. D. Burkart, *ChemBioChem*, 2015, **16**, 528–547.

118 J. A. Shields, A. S. Rahman, C. J. Arthur, J. Crosby, J. Hothersall, T. J. Simpson and C. M. Thomas, *ChemBioChem*, 2010, **11**, 248–255.

119 J. Crosby, D. H. Sherman, M. J. Bibb, W. P. Revill, D. A. Hopwood and T. J. Simpson, *Biochim. Biophys. Acta*, 1995, **1251**, 32–42.

120 N. M. Kosa, K. M. Pham and M. D. Burkart, *Chem. Sci.*, 2014, **5**, 1179–1186.

121 H. S. T. Bukhari, R. P. Jakob and T. Maier, *Structure*, 2014, **22**, 1775–1785.

122 R. E. Williams and N. C. Bruce, *Microbiology*, 2002, **148**, 1607–1614.

123 L. M. M. Mouterde and J. D. Stewart, *Org. Process Res. Dev.*, 2016, **20**, 954–959.

124 V. Agarwal, S. Diethelm, L. Ray, N. Garg, T. Awakawa, P. C. Dorrestein and B.

S. Moore, *Org. Lett.*, 2015, **17**, 4452–4455.

125 J. Zheng, C. A. Taylor, S. K. Piasecki and A. T. Keatinge-Clay, *Structure*, 2010, **18**, 913–922.

126 D. T. Wagner, J. Zeng, C. B. Bailey, D. C. Gay, F. Yuan, H. R. Manion and A. T. Keatinge-Clay, *Structure*, 2017, **25**, 1045–1055.

127 D. C. Stevens, D. T. Wagner, H. R. Manion, B. K. Alexander and A. T. Keatinge-Clay, *J. Antibiot. (Tokyo).*, 2016, **69**, 567–570.

128 I. Nazi, K. P. Koteva and G. D. Wright, *Anal. Biochem.*, 2004, **324**, 100–105.

129 E. Strauss and T. P. Begley, *J. Biol. Chem.*, 2002, **277**, 48205–48209.

130 D. C. Gay, P. J. Spear and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2014, **9**, 2374–2381.

131 L. Zhang, J. Ji, M. Yuan, Y. Feng, L. Wang, Z. Deng, L. Bai and J. Zheng, *ACS Chem. Biol.*, 2018, **13**, 871–875.

132 A. S. Worthington and M. D. Burkart, *Org. Biomol. Chem.*, 2006, **4**, 44–46.

133 X. Dong, C. D. Bailey, C. Williams, J. Crosby, T. J. Simpson, C. L. Willis and M. P. Crump, *Chem. Sci.*, 2016, **7**, 1779–1785.

134 B. Uytterhoeven, T. Lathouwers, M. Voet, C. W. Michiels and R. Lavigne, *Front. Microbiol.*, 2016, **7**, 1726.

135 T. Ackrill, PhD Thesis, University of Bristol, 2015.

136 S. E. Evans, C. Williams, C. J. Arthur, E. Płoskoń, P. Wattana-amorn, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and M. P. Crump, *J. Mol. Biol.*, 2009, **389**, 511–528.

137 E. Płoskoń, C. J. Arthur, A. L. P. Kanari, P. Wattana-Amorn, C. Williams, J. Crosby, T. J. Simpson, C. L. Willis and M. P. Crump, *Chem. Biol.*, 2010, **17**, 776–785.

138 N. J. Greenfield, *Nat. Protoc.*, 2007, **1**, 2876–2890.

139 M. A. Skiba, F. P. Maloney, Q. Dan, A. E. Fraley, C. C. Aldrich, J. L. Smith and W. C. Brown, *Methods Enzymol.*, 2018, **604**, 45–88.

140 D. A. Herbst, R. P. Jakob, F. Zähringer and T. Maier, *Nature*, 2016, **531**, 533–537.

141 H. M. Miziorko and M. D. Lane, *J. Biol. Chem.*, 1977, **252**, 1414–1420.

142 Y. Li, E. Luxenburger and R. Müller, *Angew. Chemie Int. Ed.*, 2013, **52**, 1304–1308.

143 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. Mcgettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, 2947–2948.

144 M. M. Benning, T. Haller, J. A. Gerlt and H. M. Holden, *Biochemistry*, 2000, **39**, 4630–4639.

145 M. J. Bennett, M. P. Schlunegger and D. Eisenberg, *Protein Sci.*, 1995, **4**, 2455–2468.

146 D. C. Gay, D. T. Wagner, J. L. Meinke, C. E. Zogzas, G. R. Gay and A. T. Keatinge-Clay, *J. Struct. Biol.*, 2016, **193**, 196–205.

147 E. B. Fuller, A. T.; Banks, G.T.; Mellows, G.; Barrow, K.D.; Woolford, M.; Chain, *Nature*, 1971, **234**, 416–417.

148 E. B. Chain and G. Mellows, *J. Chem. Soc. Chem. Commun.*, 1974, **20**, 847–848.

149 E. B. Chain and G. Mellows, *J. Chem. Soc., Perkin Trans. 1*, 1977, 318–322.

150 J. P. Clayton, P. J. O'Hanlon and N. H. Rogers, *Tetrahedron Lett.*, 1980, **21**, 881–884.

151 T. C. Feline, R. B. Jones, G. Mellows and L. Phillips, *J. Chem. Soc., Perkin Trans. 1*, 1977, 309–318.

152 J. P. Clayton, R. S. Oliver, N. H. Rogers and T. J. King, *J. Chem. Soc., Perkin Trans. 1*, 1979, 838–846.

153 L. Wang, Unpublished Work.

154 J. Hughes and G. Mellows, *Biochem. J.*, 1978, **176**, 305–318.

155 L. F. Silvian, M. Ile, J. Wang and T. A. Steitz, *Science*, 1999, **285**, 1074–1077.

156 T. Nakama, O. Nureki and S. Yokoyama, *J. Biol. Chem.*, 2001, **276**, 47387–47393.

157 A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419–430.

158 S. S. Gao, J. Hothersall, J. Wu, A. C. Murphy, Z. Song, E. R. Stephens, C. M. Thomas, M. P. Crump, R. J. Cox, T. J. Simpson and C. L. Willis, *J. Am. Chem. Soc.*, 2014, **136**, 5501–5507.

159 L. Wang, A. Parnell, C. Williams, N. A. Bakar, M. R. Challand, M. W. van der Kamp, T. J. Simpson, P. R. Race, M. P. Crump and C. L. Willis, *Nat. Catal.*, 2018, **1**, 968–976.

160 R. Gurney and C. M. Thomas, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 11–21.

161 S. S. Gao, L. Wang, Z. Song, J. Hothersall, E. R. Stevens, J. Connolly, P. J. Winn, R. J. Cox, M. P. Crump, P. R. Race, C. M. Thomas, T. J. Simpson and C. L. Willis, *Angew. Chemie Int. Ed.*, 2017, **56**, 3930–3934.

162 J. Hothersall, J. Wu, A. S. Rahman, J. A. Shields, J. Haddock, N. Johnson, S. M. Cooper, E. R. Stephens, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2007, **282**, 15451–15461.

163 J. Wu, S. M. Cooper, R. J. Cox, J. Crosby, M. P. Crump, J. Hothersall, T. J. Simpson, C. M. Thomas and C. L. Willis, *Chem. Commun.*, 2007, **20**, 2040–2042.

164 S. M. Cooper, W. Laosripaiboon, A. S. Rahman, J. Hothersall, A. K. El-Sayed, C. Winfield, J. Crosby, R. J. Cox, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2005, **12**, 825–833.

165 C. M. Thomas, J. Hothersall, C. L. Willis and T. J. Simpson, *Nat. Rev. Microbiol.*, 2010, **8**, 281–289.

166 X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, G. Gottschalk, R. D. Sussmuth and R. Borriss, *J. Bacteriol.*, 2006, **188**, 4024–4036.

167 W. Zhang, B. D. Ames, S. C. Tsai and Y. Tang, *Appl. Environ. Microbiol.*, 2006, **72**, 2573–2580.

168 S. C. Findlow, C. Winsor, T. J. Simpson, J. Crosby and M. P. Crump, *Biochemistry*, 2003, **42**, 8423–8433.

169 J. Wu, J. Hothersall, C. Mazzetti, Y. O'Connell, J. A. Shields, A. S. Rahman, R. J. Cox, J. Crosby, T. J. Simpson, C. M. Thomas and C. L. Willis, *ChemBioChem*, 2008, **9**, 1500–1508.

170 J. Hothersall, A. C. Murphy, Z. Iqbal, G. Campbell, E. R. Stephens, J. Wu, H. Cooper, S. Atkinson, P. Williams, J. Crosby, C. L. Willis, R. J. Cox, T. J. Simpson and C. M. Thomas, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 1017–1026.

171 A. C. Murphy, S.-S. Gao, L.-C. Han, S. Carobene, D. Fukuda, Z. Song, J. Hothersall, R. J. Cox, J. Crosby, M. P. Crump, C. M. Thomas, C. L. Willis and T. J. Simpson, *Chem. Sci.*, 2014, **5**, 397–402.

172 B. Li, W. J. Wever, C. T. Walsh and A. A. Bowers, *Nat. Prod. Rep.*, 2014, **31**, 905–923.

173 Z. Qin, S. Huang, Y. Yu and H. Deng, *Mar. Drugs*, 2013, **11**, 3970–3997.

174 I. A. Berg, D. Kockelkorn, W. Buckel and G. Fuchs, *Science (80-.).*, 2007, **318**, 1782–1786.

175 D. Fukuda, A. S. Haines, Z. Song, A. C. Murphy, J. Hothersall, E. R. Stephens, R. Gurney, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and C. M. Thomas, *PLoS One*, 2011, **6**, e18031.

176 A. M. Gulick, *ACS Chem. Biol.*, 2009, **4**, 811–827.

177 S. Schmelz and J. H. Naismith, *Curr. Opin. Struct. Biol.*, 2009, **19**, 666–671.

178 R. D. Süssmuth and A. Mainz, *Angew. Chemie Int. Ed.*, 2017, **56**, 3770–3821.

179 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505.

180 R. B. Hamed, L. Henry, J. R. Gomez-Castellanos, A. Asghar, J. Brem, T. D. W. Claridge and C. J. Schofield, *Org. Biomol. Chem.*, 2013, **11**, 8191–8196.

181 T. L. Hwang and A. J. Shaka, *J. Magn. Reson. Ser. A*, 1995, **112**, 275–279.

182 R. D. Peterson, C. A. Theimer, H. Wu and J. Feigon, *J. Biomol. NMR*, 2004, **28**, 59–67.

183 A. L. Breeze, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2000, **36**, 323–372.

184 C. Zwahlen, P. Legault, S. J. F. Vincent, J. Greenblatt, R. Konrat and L. E. Kay, *J.*

Am. Chem. Soc., 1997, **119**, 6711–6721.

185 N. M. Gaudelli and C. A. Townsend, *J. Org. Chem.*, 2013, **78**, 6412–6426.

186 A. Verma, D. M. Wong, R. Islam, F. Tong, M. Ghavami, J. M. Mutunga, C. Slebodnick, J. Li, E. Viayna, P. C. H. Lam, M. M. Totrov, J. R. Bloomquist and P. R. Carlier, *Bioorganic Med. Chem.*, 2015, **23**, 1321–1340.

187 S. Mo, D. H. Kim, J. H. Lee, J. W. Park, D. B. Basnet, Y. H. Ban, Y. J. Yoo, S. W. Chen, S. R. Park, E. A. Choi, E. Kim, Y. Y. Jin, S. K. Lee, J. Y. Park, Y. Liu, M. O. Lee, K. S. Lee, S. J. Kim, D. Kim, B. C. Park, S. G. Lee, H. J. Kwon, J. W. Suh, B. S. Moore, S. K. Lim and Y. J. Yoon, *J. Am. Chem. Soc.*, 2011, **133**, 976–985.

188 E. Bardshiri, T. J. Simpson, A. Ian Scott and K. Shishido, *J. Chem. Soc. Perkin Trans. 1*, 1984, 1765–1767.

189 H. A. Duong, P. B. Huleatt, Q. W. Tan and E. L. Shuying, *Org. Lett.*, 2013, **15**, 4034–4037.

190 D. Seebach and M. F. Züger, *Tetrahedron Lett.*, 1984, **25**, 2747–2750.

191 K. Mori and T. Ebata, *Tetrahedron*, 1986, **42**, 4413–4420.

192 J. C. Barrish, H. L. Lee, T. Mitt, G. Pizzolato, E. G. Baggiolini and M. R. Uskokovic, *J. Org. Chem.*, 1988, **53**, 4282–4295.

193 C. Bauder, *Tetrahedron Lett.*, 2008, **49**, 2243–2246.

194 P. Renaud and D. Seebach, *Helv. Chim. Acta*, 1986, **69**, 1704–1710.

195 S. Nishiguchi, M. O. Sydnes, A. Taguchi, T. Regnier, T. Kajimoto, M. Node, Y. Yamazaki, F. Yakushiji, Y. Kiso and Y. Hayashi, *Tetrahedron*, 2010, **66**, 314–320.

196 S. G. Davies, O. Ichihara, P. M. Roberts and J. E. Thomson, *Tetrahedron*, 2011, **67**, 216–227.

197 J. Zhu and D. Ma, *Angew. Chemie Int. Ed.*, 2003, **42**, 5348–5351.

198 A. D. Fotiadou and A. L. Zografos, *Org. Lett.*, 2011, **13**, 4592–4595.

199 M. R. Harris, M. O. Konev and E. R. Jarvo, *J. Am. Chem. Soc.*, 2014, **136**, 7825–7828.

200 P. A. Ledin, F. Friscourt, J. Guo and G. J. Boons, *Chem. - A Eur. J.*, 2011, **17**, 839–846.

201 V. V Vintonyak and M. E. Maier, *Org. Lett.*, 2007, **9**, 655–658.

5 Experimental

5.1 Materials and Methods

Reagents were purchased from Sigma Aldrich, Thermo Fisher or Merck Millipore. Isotopically enriched $^{15}NH_4Cl$ and $^{13}C_6$-D-glucose were purchased from Goss Scientific. *E. coli* competent cells were purchased from New England Biolabs (T7 express and 5-α) or Merck Millipore (Novagen BL21 (DE3)). All enzymes used were purchased from Thermo Fisher Scientific.

Buffers were prepared by adding the solid reagents and made up to 80% of the final volume using distilled water. The pH was adjusted using either 6 M HCl or 6 M NaOH and monitored using a Jenway pH meter. Buffers required for protein purification using a GE Healthcare AKTA FPLC were vacuum filtered through a Whatman cellulose nitrate 0.22 µM membrane and degassed by stirring under reduced pressure for 15 minutes. Buffers were stored at 4 °C and warmed to room temperature prior to use. Antibiotic and glucose solutions were sterilised with a 0.22 µM syringe filter (Thermo)

Recipes

LB broth	Tryptone (10 g/L), yeast extract (5 g/L) and NaCl (5 g/L)
LB agar	Tryptone (10 g/L), yeast extract (5 g/L) and NaCl (5 g/L), agar (15 g/L)
M9	Na_2HPO_4 (13.6 g/L), KH_2PO_4 (6 g/L), NaCl (1 g/L), pH 7.0
Re-suspension Buffer	50 mM Tris, 0.5 M NaCl, 10% (v/v) glycerol, 0.01% (v/v), Triton X-100, pH 8.0
Column Buffer A	50 mM Tris, 0.5 M NaCl, 10% (v/v) glycerol, pH 8.0
Column Buffer B	50 mM Tris, 0.1 M NaCl, 0.8 M imidazole, 10% (v/v) glycerol, pH 8.0
Reaction Buffer	50 mM Tris, 0.1 M NaCl, pH 8.0
Storage Buffer	50 mM Tris, 0.1 M NaCl, 10% (v/v) glycerol, pH 8.0
NMR Buffer	50 mM Sodium Phosphate, 0.1 M NaCl, 10% D_2O, pH 8.0 (10% D_2O was added prior to transferring to an NMR tube).
Trace metal mix (1000x) stock	$FeCl_3.6H_2O$ (50 mM), $CaCl_2$ (20 mM), $MnCl_2.4H_2O$ (10 mM), $ZnSO_4.7H_2O$ (2 mM), $CoCl_2.6H_2O$ (2 mM), $CuCl_2.2H_2O$ (2 mM), $NiCl_2.6H_2O$ (2 mM), $Na_2MoO_4.2H_2O$ (2 mM), $Na_2SeO_3.5H_2O$ (2 mM), H_3BO_3, HCl (60 mM)

Sterile Technique

All media and glassware were autoclaved prior to use to minimise contamination. An Astell Scientific ASA270 autoclave ran at 120 °C, 15 psi for 15 minutes and media was cooled to room temperature before use. Procedures requiring a sterile environment were either carried out under aseptic conditions with a Bunsen flame or in a Medical Air Technology LTD BioMat Class II microbiological safety cabinet.

PCR

Primers were designed in house then synthesised by IPT Technologies or Sigma Aldrich. Forward and reverse primers (bold type) incorporated an adapter region (normal type) to allow recombination into a pre-linearised POPINF vector. 100 μM primer stock solutions were prepared in sterile water and 25 μM working solutions were prepared by further dilution. PCR reactions were performed using a KOD DNA Polymerase kit (Merck Millipore) or KAPA HiFi HotStart ReadyMixPCR kit (KAPABIOSYSTEMS) following manufacturer's instructions. Gradient PCR was run to optimise annealing temperature for each primer pair. All reactions were performed using a peqSTAR 96x Universal Gradient or a SureCycler 8800 (Agilent Technologies)and 200 μL PCR tubes. To confirm amplification, the products were analysed by gel electrophoresis using a 1% agarose gel. PCR products were purified using a PureLink PCR Purification kit (Thermo Fisher).

KOD DNA Polymerase kit

Component	Volume (μL)
10x Buffer	10
$MgSO_4$	6
dNTPs	10
Forward Primer	3
Reverse Primer	3
Template DNA	2
Water	64
KOD Hot Start DNA Polymerase	2
Total Volume	100

1) Polymerase activation at 95 °C for 5 mins; 2) Denaturation at 95 °C for 45 s; 3) Annealing at optimal temperature for 30 s; 4) Extension for 30 s. Steps 2-4 were repeated 30 times prior to a final extension at 70 °C for 10 mins.

KAPA HiFi HotStart ReadyMixPCR kit

Component	Volume (µL)
2 x KAPA premix	25
Forward Primer	1.5
Reverse Primer	1.5
Template DNA	1
Water	21
Total Volume	50

1) Initial denaturation at 95 °C for 3 mins; 2) Denaturation at 98 °C for 20 s; 3) Annealing at optimal temperature for 15 s; 4) Extension at 72 °C for 30 s. Step 2-4 were repeated 30 times prior to a final extension at 72 °C for 10 mins.

Amplification Primers

Primer	Sequence
ACP4_For	aagttctgtttcagggcccg**AGCCAGGTTAAAGAAGGTC**
ACP4_Rev	atggtctagaaagctttta**TGCCAGCTCGCTTTCTTTAA**
ACP5_For	aagttctgtttcagggcccg**TCTGTTGCTGACTTGCGAATGGAGTTGG**
ACP5_Rev	atggtctagaaagctttta**TCAACAAATGGCTCGCCAACACTTTGATAG**
ACP6_For	aagttctgtttcagggcccg**GATCTCTCAAAGCTGGAGGCAGAGCTCTGT**
ACP6_Rev	atggtctagaaagctttta**CCGCCAGGAACAGAGTGAAGCGCTTGAG**
4M_For	aagttctgtttcagggcccg**ATGAGTCAGGTCAAAGAGGGGTTGCGAAGA**
4M_Rev	atggtctagaaagctttta**TTAATGGTCATGCAGCGCGTCCAGCC**
Short_mECH_For	aagttctgtttcagggcccg**CGCGAGTCAGTGGTCC**
Short_mECH_Rev	atggtctagaaagctttta**TTAGTTGTCGACAAAGGTTT**
BatK_For	aagttctgtttcagggcccg**ATGATGCCTATGGCTACCATTACGGC**
BatK_Rev	atggtctagaaagctttta**TTACGCCAGATAACCCGACATGGACACAAA**
TacpA_For	aagttctgtttcagggcccg**ATGGAAGGAATATACCAATTTGTTC**
TacpA_Rev	atggtctagaaagctttta**CTAAGCAACTTCACCTTCAACT**
TacpD_For	aagttctgtttcagggcccg**ATGTTAAACGTACAGCAAATTAAGAATGAA**
TacpD_Rev	atggtctagaaagctttta**TCACCTTACTTGGTTTTTAATGGCCTTAAG**
TmlQ_For	aagttctgtttcagggcccg**AATAGAACATGGATTTCCGATCGAGTAAAC**
TmlQ_Rev	atggtctagaaagctttta**TTATATCTTATTGATTTTTTTGAATCTTTC**
TmlS_For	aagttctgtttcagggcccg**ATGGAAAGTGAAAATGGTACAGTGATAATA**
TmlS_Rev	atggtctagaaagctttta**TTAACATGTAAAACCTCCGTCTATTACGAT**
MacpA_For	aagttctgtttcagggcccg**ATGAACCCTGAAAGGCGGAA**
MacpA_Rev	atggtctagaaagctttta**TTATCATGCCGAGGCTGCCG**
MupQ_For	aagttctgtttcagggcccg**ATGAGAGAGGAACGTAATTGGATCAGC**
MupQ_Rev	atggtctagaaagctttta**TCATTGCGAGGCCACCTTG**
MupS_For	aagttctgtttcagggcccg**ATGACTGATGCAGTTTCTGACGCTTTA**
MupS_Rev	atggtctagaaagctttta**TCAGCATGTGAAACCTCCATCGACT**

Mutagenic Primers

Primer	Sequence
ACP3_For	GAAACCGAACTGGCA**TAA**CGTGCACAGCGTACC
ACP3_Rev	GGTACGCTGTGCACG**TTA**TGCCAGTTCGGTTTC

Pre-linearising POPINF

16 µL of POPINF plasmid (~100 ng/µL) was incubated with 10x Fast Digest buffer (2 µL), 10 units/µL KpnI (1 µL) and 20 units/µL HindIII (1 µL) and incubated at 37 °C for 3 hours. The DNA was purified by PCR purification kit (Thermo).

Cloning

PCR fragments were recombined into a pre-linearised POPINF vector using either T4 polymerase or Infusion (Clontech).

For T4 polymerase, the reagents were mixed and incubated at room temperature for 15 mins.

Reagent	Volume (µL)
Water	4
PCR product	2
Cut POPINF vector	1
10x buffer	1
BSA	1
T4 DNA polymerase	1

For Infusion reactions, the reagents were mixed and incubated at 50 °C for 20 mins.

Reagent	Volume (µL)
Water	4
PCR product	3
Cut POPINF vector	1
5 x Infusion premix	1

Following incubation, 3 μL of the reaction was transformed into 50 μL NEB 5-α *E. coli* competent cells and plated on LB agar plates supplemented with carbenicillin (100 μg/mL), X-gal (40 μg/mL) and IPTG (1 mM). Single, white colonies were picked, cultured in LB (5 mL) supplemented with carbenicillin (100 μg/mL) for 16 h at 37 °C. The cells were pelleted by centrifugation at 6,000 rpm for 5 mins and the plasmid DNA purified using a GenElute (Sigma Aldrich) or QIAprep (Qiagen) Spin Miniprep Kit. The plasmid DNA was sequenced by Genewiz.

Plasmids

The nucleotide sequence for BatA, BatD, BatE, ACP3/4, mECH and MacpD were codon optimised for *E. coli*, then synthesised and subcloned into a pET151/D-TOPO vector by Thermo Fisher.

Transformation

To 20 μL of *E. coli* competent cells thawed on ice was added 1 μL of plasmid DNA. The cells were incubated on ice for 30 min then heat shocked by incubating at 42 °C for 30 s and cooled on ice for 2 mins. 200 μL SOC medium was added and the mixture incubated at 37 °C with shaking at 220 rpm for 1 h. The cells were plated, diluted where appropriate with SOC, on to LB agar supplemented with antibiotic (100 μg/mL) and incubated at 37 °C for 16 h.

General Expression Procedure

E. coli competent cells transformed with plasmid DNA were grown on LB agar supplemented with 100 μg/mL carbenicillin (LB/carb) at 37 °C overnight. A single colony was picked and added to 50 mL LB/carb and grown at 37 °C for 16 h with shaking at 200 rpm. 2 mL of this seed broth was used to inoculate flasks containing 200 mL LB/carb medium and incubated at 37 °C with shaking at 200 rpm until an OD_{600} of 1.0 was reached. The culture was then induced with 250 μM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated for 16 h at 16 °C. The cells were harvested by centrifugation at 6,000 rpm and re-suspended in column buffer A (50 mM Tris, 500 mM NaCl, 10% glycerol, pH 8.0). The cells were lysed by sonication and the soluble fraction applied to a HisTrap HP 5 mL Ni column (GE Life Sciences). Protein was eluted *via* a linear gradient from 5 to 100% column buffer B (50 mM Tris, 100 mM NaCl, 10% glycerol, 0.8 M imidazole, pH 8.0) and collected in 1 mL fractions. Fractions were analysed by SDS-

PAGE and those containing the protein of interest were pooled, buffer exchanged in to reaction buffer (50 mM Tris, 100 mM NaCl, pH 8.0) using a HiPrep 26/10 desalting column (GE Life Sciences) or further purified by Superdex S75 size exclusion chromatography (SEC) using reaction buffer. Protein was concentrated and stored at 4 °C, or supplemented with 10% glycerol, aliquoted, flash frozen and stored at -80 °C.

M9 expression

A single colony was picked and cultured in 50 mL LB medium supplemented with carbenicillin (100 µg/mL) for 16 h at 37 °C. 2 mL of pre-culture was added to 100 mL M9 containing carbenicillin (100 µg/mL), NH_4Cl (1 g/L), glucose (2 g/L) and metal mix (100 µL) and incubated at 37 °C until an optical density of 1.0 was reached. The culture was induced with IPTG (250 µM) and incubated for 16 h at 16 °C. The protein was purified as per the unlabelled material.

M9 expression *via* media transfer

A single colony was picked and cultured in 50 mL LB medium supplemented with carbenicillin (100 µg/mL) for 16 h at 37 °C. 2 mL of pre-culture was added to 200 mL LB/carb and incubated at 37 °C until OD_{600} 1.0 was reached. The cells were harvested by centrifugation (6000 rpm for 10 mins) and washed twice with sterile M9 medium. They were then resuspended in M9 and used to inoculate 100 mL M9 supplemented with carbenicillin (100 µg/mL), NH_4Cl (1 g/L), glucose (2 g/L) and metal mix (100 µL) and incubated at 16 °C for 1 h. The culture was induced with IPTG (250 µM) and incubated for 16 h at 16 °C. The protein was purified as per the unlabelled material.

Expression of ACPs

ACP4 and 4M were expressed following the general procedure, except 1 mM IPTG was used to induce the culture, which was then incubated at 16 °C for 5 h. All modular ACPs (3/4, 4, 5, 6 and 4M) were desalted into reaction buffer (~100 µM) and were stable at 4 °C for up to 7 days. ACP3/4, ACP4 and ACP5 were stable to freeze-thaw conditions, ACP6 was not stable to these conditions and 4M was not frozen.

BatC

BatC was expressed and purified following the general procedure, desalted into reaction buffer supplemented with 10% glycerol and flash frozen in 50 µL aliquots (300 µM).

BatD and BatE

BatD and BatE were expressed and purified following the general procedure. Both proteins (~200 µM) could be stored at 4 °C for up to 3 days, however, loss of activity was observed after 1 week at 4 °C. Both proteins were stable to a single freeze-thaw cycle, however, catalytic activity declined with number of freeze-thaw cycles and the length of storage. Optimal results were obtained with freshly purified protein.

MupQ and TmlQ

MupQ and TmlQ were expressed and purified following the general procedure and desalted into reaction buffer supplemented with 10% glycerol. Difficulties concentrating both proteins were encountered due to precipitation in a Vivaspin (GE Life Sciences) or Amicon (Merck) spin concentrator with a 10 kDa molecule weight cut-off. A maximum concentration of ~50 µM was achieved and the protein was flash frozen in aliquots.

MupS and TmlS

MupS and TmlS were both expressed according to the general procedure. During purification, 0.5 M NaCl and 10% glycerol was required in the reaction buffer to prevent precipitation on the desalting column.

Holo-ACP preparation

The reagents were added to apo-ACP in reaction buffer and incubated at room temperature for 1 h. Reaction progress was monitored by ESI-MS.

Reagent	Stock	Volume (µL)	Final
apo-ACP		87	50-100 µM
$MgCl_2$	1 M	1	10 mM
CoA	100 mM	1	1 mM
MupN	15 µM	10	1.5 µM
TCEP	100 mM	1	1 mM

One-pot pantetheine upgrade and loading of ACPs

Adapting literature procedures,[124,180] to apo-ACP in reaction buffer was added each reagent in series and the reaction mixture was incubated at room temperature until complete loading was observed. Reaction progress was monitored by ESI-MS and typical

reaction times were 1 h (100 µL reactions) or 1-3 h (1 mL or larger reactions). Once full loading was achieved, excess reagents were removed using a HiPrep 26/10 desalting column (GE Life Sciences) or Zeba column (Thermo Fisher) equilibrated with reaction buffer. Loaded ACPs were concentrated using Sartorius Vivaspin20 with a 3 or 10 kDa molecular weight cut-off.

Reagent	Stock	Small-scale Volume (µL)	Large-scale Volume (µL)	Final Concentration
ACP		84	900	50-100 µM
$MgCl_2$	1 M	1	10	10 mM
Pantetheine	100 mM	1	5	1/0.5 mM
CoA mix (A, D, E)	33 µM each	3	30	1 µM each
MupN	15 µM	10	100	1.5 µM
ATP	100 mM	1	10	1 mM

ESI-MS sample preparation

Samples were desalted for ESI-MS analysis using a C_4 ZipTip™ (Millipore). The Ziptip was activated with two 10 µL aspirations of solution A (50% acetonitrile in water) and equilibrated with five 10 µL aspirations of solution B (0.1% TFA in water). A 10 µL sample was then loaded onto the ZipTip with 15-20 aspirations, returning the sample to the same tube between aspirations. The protein was desalted with ten 10 µL washes with solution C (5% MeOH, 0.1% TFA in water) and discarded to waste. Finally, the sample was eluted with 10 µL of solution D (70% acetonitrile, 0.1% TFA in water).

Denatured samples were analysed on a Synapt G2-Si (Waters) fitted with a TriVersa NanoMate® (Advion) using the following parameters: sample cone 10 V, capillary voltage 1.5 kV, transfer collision energy 5 V. The source was set to positive mode and spectra were acquired over 600-3000 m/z and analysed using MassLynx™ 4.1 software. The spectra were smoothed, and background subtraction applied. Deconvolution of the charge states was undertaken using MagTran 1.03. For Ppant ejection assays, an appropriate charge state was isolated using the MSMS functionality. The trap collision energy was increased until fragmentation was observed (typically between 5 V and 30 V) and spectra were collected from 200-2000 *m/z*.

For analysis by LC/MS, assay samples were diluted with acetonitrile (v/v) and 20 µL injected for analysis. An UltiMate 3000 UPLC (Thermo) fitted with an ACQUITY UPLC Protein BEH C4 column (300 Å, 1.7 µM, 2.1 mm x 150 mm) was used for LC separation. HPLC grade H_2O (**A**) and acetonitrile (**B**) supplemented with 0.1 % formic acid were used as the solvents at a flow rate of 0.2 mL/min. The following gradient was run:

Time	Eluent
0-2 min	30% B
2-22 min	Linear gradient to 30-50% B
22-23 min	50% B
23-24 min	Linear gradient to 30% B
24-25 min	30% B

The sample was analysed using an Orbitrap Elite (Thermo) across the mass range 200-2000 Da. Collison-induced dissociation (CID) at 20 V was used for Ppant ejection assays.

β-branching assay by MS

BatA was loaded with Ac-pantetheine **60** and the modular ACP (ACP4, 5 or 6) was loaded with Acac-pantethine **62** (100 µL total volume per reaction). Once full loading was achieved, the sample was desalted into reaction buffer using a Zeba column (Thermo Fisher). Ac-BatA (50 µM) and Acac-ACP (50 µM) were mixed and BatC (5 µM) added to the 100 µL reaction, which was incubated at room temperature for 1 h. A 10 µL sample was taken for MS analysis and 5 µM of the required catalytic domains (BatD/BatE/mECH) was added. The reaction was incubated at room temperature for 1 h and analysed by MS (10 µL aliquot).

BatK reduction

The modular ACP (~100 µM) was loaded with enoyl-pantetheine **64** and monitored by MS until complete loading was achieved (~1 h). BatK (5 µM) and NADH (1 mM) were added and the reaction mixture incubated at room temperature for 1 h. MS analysis by Ppant ejection was used to determine the outcome of the reaction.

β-branching assay by NMR

Apo-BatA was loaded with [^{13}C]-Ac-pantethine **85**, monitored by MS until complete loading was achieved (typically 1 to 3 h). The loaded ACP was desalted into NMR buffer using a HiPrep 26/10 desalting column and concentrated using Vivaspin (GE Life

Sciences) or Amicon (Merck) spin concentrators. [^{13}C]-Ac-BatA was stored in aliquots at -80 °C and single aliquots thawed on ice when required. Modular ACPs were freshly loaded with Acac-pantethine **62**, desalted and concentrated as described, however, these ACPs would be used immediately due to hydrolytic instability of the β-keto thioester.

To form [^{13}C]-HMG-ACP, Acac-ACP, [^{13}C]-Ac-BatA and BatC were incubated at room temperature for 1 h. 10% D_2O was added to the sample and the mixture transferred to an NMR tube for analysis. To assess ECH function, [^{13}C]-HMG-ACP was formed over 1 h, then BatD and/or BatE were added to the reaction. After incubation for 1 h, 10% D_2O was added and the sample transferred to an NMR rube.

Optimisation meant the concentration of modular ACP, BatA and the catalytic enzymes varied with each experiment according to the table below. This was determined by protein expression, stock concentrations and dilution during NMR sample preparation.

ACP_A	**ACP_A (μM)**	**ACP_D**	**ACP_D (μM)**	**ACP_A:ACP_D**	**BatC (μM)**	**BatD (μM)**	**BatE (μM)**
[^{13}C]-Ac-BatA	345						
[^{13}C]-Ac-ACP4	340						
Acac-ACP4	215	[^{13}C]-Ac-BatA	198	1.1	7		
Acac-ACP4	342	[^{13}C]-Ac-BatA	145	2.4	5.4	26.8	
Acac-ACP4	267	[^{13}C]-Ac-BatA	196	1.4	7	5	5
Acac-ACP5	156	[^{13}C]-Ac-BatA	168	0.9	13.5	10.2	9.8
[^{13}C]-Ac-4M	150						
Acac-4M	202	[^{13}C]-Ac-BatA	150	1.3	13		
Acac-4M	226	[^{13}C]-Ac-BatA	164	1.4	5.4	24.2	
Acac-4M	261	[^{13}C]-Ac-BatA	272	1.0	10.9	13.5	11.8

NMR parameters

All protein NMR experiments were acquired on a Varian VNMRS 600 cryo spectrometer at 20 °C and spectra referenced using 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Spectra recorded in 10% D_2O used a ^{1}H DPFGSE with a sweep width of 8000 Hz with 32000 points.[181] Standard experiments were used for ^{13}C, DEPT, HSQC and HMBC TOCSY and NOESY data collection. ^{13}C,^{15}N (F1, F2) filtered NOESY and TOCSY spectra were acquired using the CNfilter-noesy and CNfilter-tocsy pulse sequences as

implemented in Biopack (VnmrJ v4.2) with calibrated ^{13}C and ^{15}N pulse widths and a ^{1}H observe sweep width of up to 8000 Hz with 2048 points.[182,183] These sequences are modifications of the original implementation by Zwahlen et al.[184]

Circular Dichroism

Samples were prepared in 50 mM sodium phosphate (pH 8.0) and the concentration adjusted to 10 µM. Spectra were averaged over 5 scans recorded at 5 °C using a 1 cm quartz cuvette and a JASCO J-1500 Spectrophotometer. Background subtraction and analysis took place using the Spectra Manager™ suite (JASCO).

Analytical S200 Gel Filtration

500 µL of protein was injected onto a Superdex 200 column (GE Healthcare) equilibrated with 25 mM Tris, 150 mM NaCl, pH7.5 at 0.5 mL/min. The column was pre-calibrated with blue dextran (2000 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), conalbumin (75 kDa), bovine serum albumin (66 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa) and cytochrome C (12.4 kDa) from a Gel Filtration Calibration Kit (GE Healthcare). By plotting the gel-phase distribution coefficient (Kav) against the logarithm of the molecular weight (Log MW) and fitting with a straight line ($y = -0.256x + 0.870$, $R^2 = 0.9876$) a calibration curve may be obtained. The elution volume of an analyte may be used to estimate its molecular weight.

$$\mathrm{Kav} = \frac{\mathrm{Ve} - \mathrm{Vo}}{\mathrm{Vc} - \mathrm{Vo}}$$

Kav = gel-phase distribution coefficient, Ve = elution volume, Vo = column void volume = 7.08 mL (based on blue dextran), Vc = geometric column volume = 24 mL.

Culturing *P. fluorescens* BCCM_ID9359

A glycerol stock of *P. fluorescens* BCCM_ID9359 was streaked onto an agar plate an incubated at 28 °C for 16 h. A single colony was picked and used to inoculate 100 mL LB medium, which was incubated at 25 °C for 16 h. The seed culture (5%) was used to inoculate 2 L LB medium supplemented with 4% glucose and incubated at 16 °C for 96 h. To harvest, the cells were removed by centrifugation (8000 rpm) and the supernatant extracted with ethyl acetate (2 x v/v), washed with sat. aq. NaCl, dried over $MgSO_4$ and concentrated *in vacuo*. The crude yellow oil was purified by two column chromatography

steps - 5% MeOH/CH_2Cl_2/0.1% AcOH followed by 2% MeOH/EtOAc/0.1% AcOH – furnishing kalimantacin A as a colourless oil (batch 1 - 38 mg/L, batch 2 - 50 mg/L).

Acidification

P. fluorescens was cultured as described above and following centrifugation the supernatant divided into two portions. One portion was acidified to pH 2 with conc. HCl, then both portions were separately extracted with ethyl acetate (2 x v/v), concentrated and purified by the standard column chromatography conditions. The non-acidified portion furnished 18 mg and the acidified portion 19 mg per 500 mL of culture.

Analytical LCMS

For analysis of natural product crude extracts, a 10 mg/mL sample was made up in HPLC-grade MeOH. 20 µL of the extracts were injected and analysed on a Waters 2795HT or Waters 2545SFO HPLC system with the following detectors:

Waters 2795HT HPLC system

- Waters 998 diode array detector for UV between 200 and 400 nm.
- Electrospray (ES) Waters ZQ mass spectrometry with detection between 150 and 800 m/z units in positive and negative modes.

Waters 2445SFO HPLC system

- Waters 2298 diode array detector for UV between 200 and 400 nm.
- Waters Quattro Micro ESI mass spectrometer in positive and negative modes, with detection between 150 and 1200 m/z units.

Both systems were equipped with a Phenomenex Kinetex column (2.6 µ, C18, 100 Å, 4.6 x 100 mm) fitted with a Phenomenex Security Guard column (Luna C5 300 Å) eluting with a flow rate of 1 mL/min. HPLC grade H_2O (**A**) and acetonitrile (**B**) supplemented with 0.05 % formic acid were used as the solvents. The following gradient was run:

Time	Eluent
0-2 min	5% B
2-15 min	Linear gradient to 90% B
15-17 min	95% B
17-18 min	Linear gradient to 5% B
18-20 min	5% B

Preparative HPLC

Preparative purification of compounds was carried out using a Waters 2545SFO HPLC system with the following set up:

- Waters 2767 autosampler
- Waters 2545 pump system.
- Phenomenex Kinetex column (5 µ, C18, 100 Å, 250 x 21.20 mm) fitted with a Phenomenex Security Guard column (Luna C5 300 Å) eluting at a flow rate of 16 mL/min.

HPLC grade H_2O (**A**) and acetonitrile (**B**) supplemented with 0.05 % formic acid were used as the solvents. The post-column flow was split (100:1) and the minor flow was made up with solvent A to 1 ml/min for simultaneous analysis by diode array detector (Waters 2998), and ESI mass spectrometry in positive and negative modes (Waters Quatro Micro). The following gradient was run:

Time	Eluent
0-2 min	5% B
2-15 min	Linear gradient to 90% B
15-16 min	95% B
16-17 min	Linear gradient to 5% B
17-20 min	5% B

Starter unit assay

Apo-ACP (~200 µM) was converted to holo-ACP by the addition of $MgCl_2$ (10 mM), CoA (1 mM), MupN (1.5 µM) and TCEP (1 mM). The reaction was monitored by MS and upon completion was desalted into reaction buffer supplemented with 1 mM TCEP using a HiPrep 26/10 desalting column. Holo-ACP was adjusted to ~100 µM, aliquoted and stored at – 20 °C. To holo-ACP (100 µM) were added MupQ/TmlQ (5 µM), malonyl/succinyl CoA (1 mM) and ATP (1 mM). The reaction was incubated at room temperature for 1 h then monitored by MS. MupS/TmlS (5 µM) and NADH (1 mM) were added and the reaction incubated for a further 1 h prior to MS analysis.

To pre-formed 3-HP-MacpD was added MupB (5 µM) and 3,3-dimethyl acrylic acid **68** or enoyl-pantetheine **64** (1 mM). The reaction was incubated at room temperature for 1 h and monitored by MS.

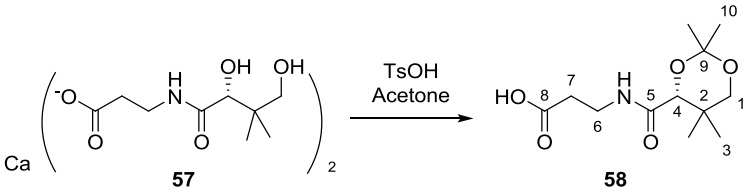
General experimental for chemical synthesis

All air and moisture sensitive reactions were carried out under an N_2 atmosphere using anhydrous solvents, obtained by passing solvent through a column of anhydrous alumina using equipment from Anhydrous Engineering and stored in a Straus flask fitted with a J. Youngs valve before use. Commercially available reagents were used without further purification. Flash column chromatography was performed with silica gel (technical grade, 40-63 μM particle size). Analytical thin layer chromatography (TLC) was performed on Merck, aluminium backed 60 F_{254} silica plates. TLC plates were visualised by UV fluorescence (UV254 lamp), basic $KMnO_4$ solution or an ethanolic solution of phosphomolybdic acid.

Nuclear magnetic resonance (NMR) spectra were recorded at ambient temperature in $CDCl_3$ (unless stated otherwise) on one of the following spectrometers; Varian 400-MR, Varian VNMRS500, Bruker Advance III HD 500 Cryo, Bruker VNMRS600 Cryo, Bruker 700 micro-cryo. ^1H NMR are reported in the format: chemical shift (ppm), integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad, dd = doublet of doublets, ddd = doublet of doublets of doublets, dt = doublet of triplets, qd = quartet of doublets), coupling constant (*J*, in Hertz), assignment. Infrared spectra were recorded on a Perkin Elmer Spectrum One Fourier Transform Infrared Spectrometer (FT-IR) with signal intensities reported as follows: br = broad, w = weak, m = medium, s= strong. Optical rotations were measured on a Bellingham and Stanley Ltd ADP220 polarimeter and melting points were determined on an Electrothermal IA6301 melting point apparatus. Small molecule high resolution mass spectrometry (HRMS) was performed on a Bruker microTOF spectrometer using electrospray ionisation (ESI).

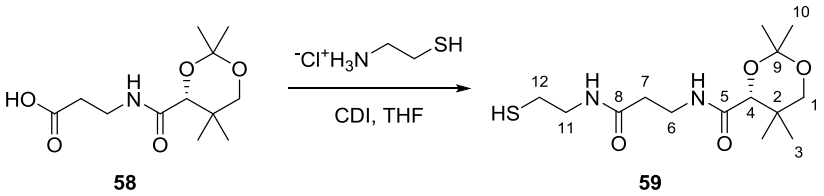
5.2 Compound Characterisation

(*R*)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propionic acid 58



To flame dried molecular sieves under a nitrogen atmosphere was added acetone (250 mL). After stirring for 20 min, D-pantothenic acid hemicalcium salt **57** (5.00 g, 21.0 mmol) and *para*-toluenesulfonic acid (4.79 g, 25.2 mmol) were added and the thick white slurry was stirred at room temperature for 16 h. The reaction mixture was filtered through a bed of Celite, washed with acetone (2 x 50 mL) and the solvent removed under reduced pressure. The residue was dissolved in EtOAc (100 mL) and washed with half saturated brine (3 x 100 mL) and dried over $MgSO_4$. The crude product was triturated with *n*-hexane furnishing **58** as a white solid (3.96 g, 73 %). $[\alpha]_D^{24}$ +62.0 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 10.50 (1H, br s, OH), 7.07 (1H, t, *J* 6.0, NH), 4.11 (1H, s, 4-H), 3.67 (1H, d, *J* 12.0, 1-*H*H), 3.63-3.53 (1H, m, 6-*H*H), 3.51-3.42 (1H, m, 6-H*H*), 3.27 (1H, d, *J* 12.0, 1-H*H*), 2.59 (2H, t, *J* 6.0, 7-H_2), 1.43 (3H, s, 10-H_3), 1.41 (3H, s, 10-H_3), 1.01 (3H, s, 3-H_3), 0.95 (3H, s, 3-H_3). δ_C (100 MHz, $CDCl_3$) 176.5 (C-8), 170.3 (C-5), 99.2 (C-9), 77.1 (C-4), 71.5 (C-1), 34.2 (C-6), 33.9 (C-7), 33.0 (C-2), 29.5 (C-10), 22.1 (C-3), 18.9 (C-3), 18.8 (C-10). ν_{max} (neat) / cm^{-1} 3388 (br, w), 2957 (w), 1723 (s), 1194 (s), 1097 (s). *m/z* (ESI^+) calc. for [$C_{12}H_{21}NO_5Na$] 282.1312 found 282.1309. The data are in accordance with the literature.[124]

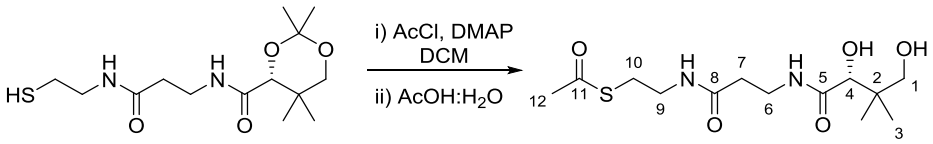
Pantetheine dimethyl ketal 59



To acid **58** (2.00 g, 7.71 mmol) in dry THF (40 mL) under a nitrogen atmosphere was added CDI (1.87 g, 11.6 mmol). The reaction mixture was stirred at room temperature for 30 min then cysteamine hydrochloride (1.31 g, 11.6 mmol) was added and stirred for 16 h. The reaction was quenched with sat. aq. NH_4Cl (30 mL) the phases separated, and the aqueous phase extracted with CH_2Cl_2 (2 x 40 mL). The combined organic phases were

washed with brine (20 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The colourless oil was purified by column chromatography (80-100% EtOAc:petrol) furnishing **59** as a white solid (1.73 g, 71%). m.p. 99-101 °C. $[\alpha]_D^{24}$ +48.0 (*c* 1.0, $CHCl_3$). Lit[185] $[\alpha]_D^{24}$ +33.4 (*c* 1.0, MeOH). δ_H (400 MHz, $CDCl_3$) 7.01 (1H, br s, NH), 6.42 (1H, br s, NH), 4.06 (1H, s, 4-H), 3.66 (1H, d, *J* 12.0, 1-*H*H), 3.61-3.49 (2H, m, 6-H_2), 3.48-3.33 (2H, m, 11-H_2), 3.26 (1H, d, *J* 12.0, 1-H*H*), 2.64 (2H, q, *J* 7.0, 12-H_2), 2.45 (2H, t, *J* 6.0, 7-H_2), 1.44 (3H, s, 10-H_3), 1.40 (3H, s, 10-H_3), 1.36 (1H, t, *J* 8.5, SH), 1.01 (3H, s, 3-H_3), 0.95 (3H, s, 3-H_3). δ_C (100 MHz, $CDCl_3$) 171.1 (C-8). 170.2 (C-5), 99.1 (C-9), 77.1 (C-4), 71.4 (C-1), 42.4 (C-11), 36.1 (C-7), 34.8 (C-6), 32.9 (C-2), 29.5 (C-10), 24.5 (C-12), 22.1 (C-3), 18.9 (C-3), 18.7 (C-10). ν_{max} (neat) / cm^{-1} 3300 (br, w), 2938 (w), 2550 (w), 1650 (s), 1524 (s), 1196 (m), 1096 (s). *m/z* (ESI^+) calc. for [$C_{14}H_{26}N_2O_4SNa$] 341.1505 found 341.1507. The data are in accordance with the literature.[124,185]

S-Acetyl-pantetheine 60

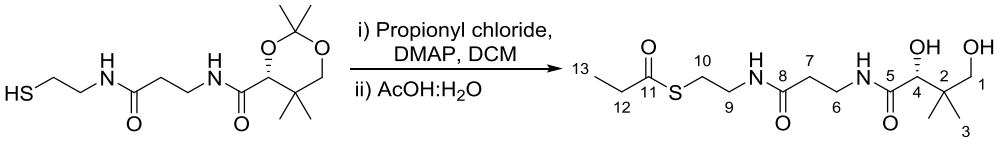


59 **60**

To thiol **59** (0.15 g, 0.47 mmol) in dry CH_2Cl_2 (10 mL) at 0 °C under a nitrogen atmosphere was added acetyl chloride (0.10 mL, 1.41 mmol) and DMAP (0.12 g, 0.94 mmol). The reaction mixture was allowed to warm to room temperature and stirred for 16 h, then quenched with 1 M HCl (15 mL). The aqueous phase was extracted with CH_2Cl_2 (2 x 20 mL), the combined organic phases were washed with brine (15 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified by column chromatography (5% MeOH:CH_2Cl_2) furnishing a colourless oil (80 mg, 47%). To the protected pantetheine (70 mg, 0.19 mmol) was added 2:1 AcOH: H_2O (8 mL), the reaction mixture stirred at room temperature for 5 h then concentrated under reduced pressure. The crude oil was purified by column chromatography (5-7% MeOH:CH_2Cl_2) furnishing **60** as a colourless oil (50 mg, 81%). $[\alpha]_D^{24}$ +22.0 (*c* 1.0, $CHCl_3$). δ_H (500 MHz, CD_3OD) 3.90 (1H, s, 4-H), 3.55-3.48 (2H, m, 6-H_2), 3.47-3.40 (2H, m, 1-H_2), 3.40-3.34 (2H, m, 9-H_2), 3.01 (2H, t, *J* 7.0, 10-H_2), 2.42 (2H, t, *J* 7.0, 7-H_2), 2.34 (3H, s, 12-H_3), 0.93 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 197.0 (C-11), 176.0 (C-5), 173.9 (C-8), 77.3 (C-4), 70.3 (C-1), 40.4 (C-2), 40.0 (C-9), 36.4 (C-7), 36.3 (C-6), 30.5 (C-12), 29.5 (C-10), 21.3

(C-3), 20.9 (C-3). ν_{max} (neat) / cm^{-1} 3298 (br), 2874 (br, w), 1644 (s), 1530 (s). *m/z* (ESI^+) calc. for [$C_{13}H_{24}N_2O_5SNa$] 343.1298 found 343.1293. The data are in accordance with the literature.[124]

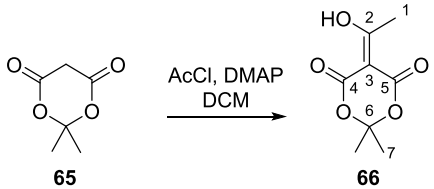
_S_-Propyl Pantetheine 61



59 **61**

To thiol **59** (0.15 g, 0.47 mmol) in dry CH_2Cl_2 (10 mL) at 0 °C under a nitrogen atmosphere was added propionyl chloride (42 µL, 0.47 mmol) and DMAP (0.12 g, 0.94 mmol). The reaction mixture was allowed to warm to room temperature and stirred for 5 h, then quenched with 1 M HCl (10 mL). The aqueous phase was extracted with CH_2Cl_2 (2 x 15 mL), the combined organic phases were washed with brine (15 mL), dried over $MgSO_4$ and concentrated *in vacuo*. To the crude protected pantetheine was added 2:1 AcOH: H_2O (9 mL), the reaction mixture stirred at room temperature for 16 h then concentrated under reduced pressure. The crude oil was purified by column chromatography (5-10% MeOH:CH_2Cl_2) furnishing **61** as a colourless oil (103 mg, 66%). δ_H (500 MHz, CD_3OD) 3.90 (1H, s, 4-H), 3.55-3.48 (2H, m, 6-H_2), 3.48-3.39 (2H, m, 1-H_2), 3.39-3.34 (2H, m, 9-H_2), 3.02 (2H, t, *J* 6.5, 10-H_2), 2.61 (2H, q, *J* 7.5, 12-CH_2), 2.42 (2H, t, *J* 6.5, 7-H_2), 1.16 (3H, t, *J* 7.5, 13-H_3), 0.94 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 201.3 (C-11), 176.0 (C-5), 173.9 (C-8), 77.3 (C-4), 70.3 (C-1), 40.4 (C-2), 40.1 (C-9), 38.2 (C-12), 36.4 (C-7), 36.3 (C-6), 29.1 (C-10), 21.3 (C-3), 20.9 (C-3), 10.0 (C-13). *m/z* (ESI^+) calc. for [$C_{14}H_{26}N_2O_5SNa$] 357.1455 found 357.1454.

Acetyl Meldrum's Acid 66

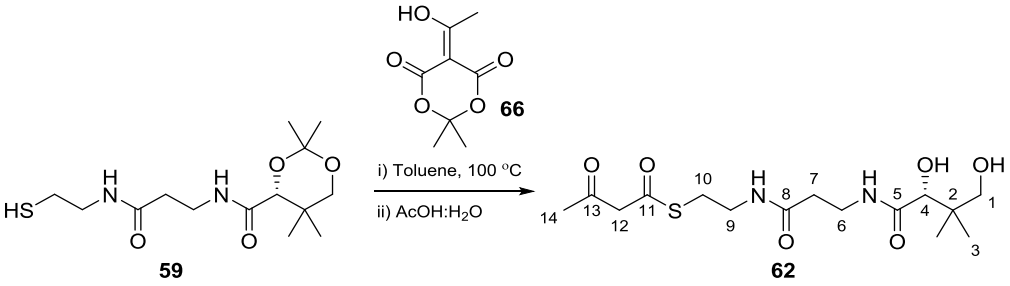


65 **66**

To Meldrum's acid **65** (0.40 g, 2.78 mmol) in dry CH_2Cl_2 (8 mL) at 0 °C under a nitrogen atmosphere was added acetyl chloride (0.24 mL, 3.33 mmol) and DMAP (0.68 g, 5.56 mmol). The reaction mixture was allowed to warm to room temperature and stirred

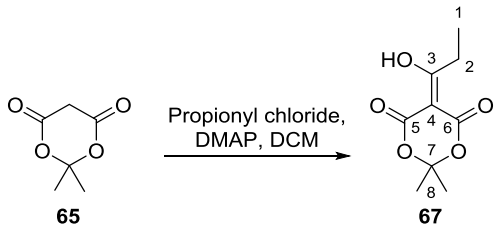
for 1 h, then quenched by the addition of 1 M HCl (15 mL). The aqueous phase was extracted with CH_2Cl_2 (2 x 20 mL), the combined organic phases washed with brine (15 mL), dried over $MgSO_4$ and concentrated *in vacuo* furnishing **66** as an orange solid (510 mg, 99%) that did not require further purification. δ_H (400 MHz, $CDCl_3$) 15.13 (1H, s, OH), 2.66 (3H, s, 1-H_3), 1.74 (6H, s, $(CH_3)_2$). δ_C (100 MHz, $CDCl_3$) 194.6 (C-2), 170.2 (CO), 160.5 (CO), 104.9 (C-6), 91.9 (C-3), 26.8 (C-7), 23.5 (C-1). The data are in accordance with the literature.[186]

_S_-Acetoacetyl-pantetheine 62



To thiol **59** (0.10 g, 0.31 mmol) and Meldrum's acid **66** (0.12 g, 0.63 mmol) under a nitrogen atmosphere was added dry toluene (4 mL). The reaction mixture was heated to reflux and stirred for 16 h. Following cooling to room temperature the solvent was removed under reduced pressure and the crude oil was purified by column chromatography (0-10% MeOH:CH_2Cl_2) furnishing a colourless oil (96 mg). To the protected pantetheine (80 mg, 0.20 mmol) was added 2:1 AcOH:H_2O (8 mL) and the reaction mixture stirred at room temperature for 5 h. The solvent was removed under reduced pressure and the crude oil was purified by column chromatography (5% MeOH:CH_2Cl_2) furnishing **62** as a colourless oil (51 mg, 54%). $[\alpha]_D^{24}$ +20.0 (*c* 1.0, MeOH). δ_H (500 MHz, CD_3OD) 3.90 (1H, s, 4-H), 3.54-3.48 (2H, m, 6-H_2), 3.48-3.41 (2H, m, 1-H_2), 3.40-3.35 (2H, m, 9-H_2), 3.07 (2H, t, *J* 6.5, 10-H_2), 2.43 (2H, t, *J* 6.5, 7-H_2), 2.24 (2.2H, s, 14-H_3 keto), 1.94 (0.6H, s, 14-H_3 enol), 0.93 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 202.4 (C-13), 193.7 (C-11), 176.0 (C-5), 174.0 (C-8), 77.3 (C-4), 70.4 (C-1), 40.4 (C-2), 39.8 (C-9), 36.4 (C-7), 36.3 (C-6), 30.3 (C-14 keto), 29.7 (C-10), 21.3 (C-3), 20.9 (C-3), 20.8 (C-14 enol). ν_{max} (neat) / cm^{-1} 3300 (br), 2931 (br, w), 1719 (w), 1644 (s), 125 (m). *m/z* (ESI^+) calc. for $[C_{15}H_{26}N_2O_6SNa]$ 385.1409 found 385.1400. The data are in accordance with the literature.[124]

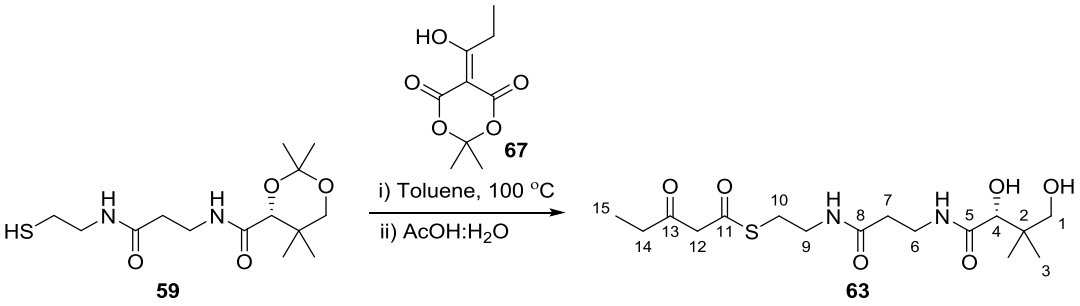
Propyl Meldrum's Acid 67



65 67

To Meldrum's acid **65** (1.00 g, 6.94 mmol) in dry CH_2Cl_2 (20 mL) at 0 °C under a nitrogen atmosphere was added propionyl chloride (0.62 mL, 6.94 mmol) in dry CH_2Cl_2 (10 mL) and DMAP (1.69 g, 13.9 mmol). The reaction mixture was allowed to warm to room temperature and stirred for 16 h, then quenched by the addition of 0.1 M HCl (40 mL). The phases were separated and the aqueous phase extracted with CH_2Cl_2 (2 x 30 mL). The combined organic phases were washed with brine (20 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified by column chromatography (80% EtOAc:petrol) furnishing **67** as an orange solid (950 mg, 68%). δ_H (400 MHz, $CDCl_3$) 3.11 (2H, q, *J* 7.5, 2-H_2), 1.73 (6H, s, $(CH_3)_2$), 1.26 (3H, t, *J* 7.5, 1-H_3). δ_H (100 MHz, $CDCl_3$) 199.1 (C-3), 170.8 (CO), 160.4 (CO), 105.0 (C-7), 91.1 (C-4), 29.6 (C2), 27.0 (C-8), 9.9 (C-1). The data are in accordance with the literature.[187]

***S*-13-oxopentanethioic-pantetheine 63**

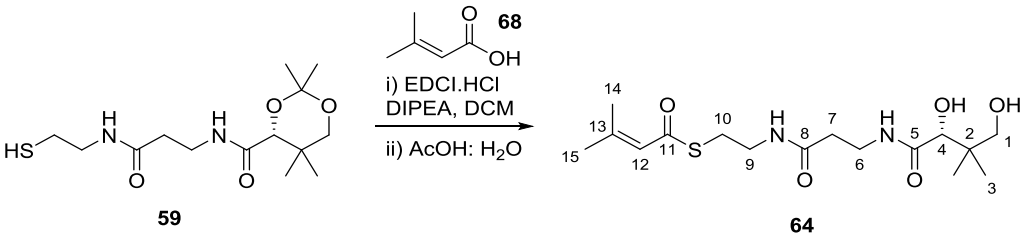


59 63

Thiol **59** (0.20 g, 0.63 mmol) and Meldrum's acid **67** (0.15 g, 0.76 mmol) in dry toluene (15 mL) under a nitrogen atmosphere were heated to 90 °C and stirred for 16 h. After cooling to room temperature, the solvent was removed under reduced pressure. To the crude orange oil was added AcOH (6 mL) and H_2O (3 mL) and the reaction mixture was stirred at room temperature for 4 h. The solvent was removed *in vacuo* and the residue was purified by column chromatography (10% MeOH:CH_2Cl_2) furnishing **63** as a light-yellow oil (75 mg, 32%). δ_H (500 MHz, CD_3OD) 3.90 (1H, s, 4-H), 3.55-3.47 (2H, m, 6-H_2), 3.47-3.40 (2H, m, 1-H_2), 3.40-3.33 (2H, m, 9-H_2), 3.07 (1.3H, t, *J* 6.5, 10-H_2), 2.83 (0.6H, t, *J* 6.5, 10-H_2), 2.60 (1.2H, q, *J* 7.0, 14-H_2), 2.48-2.40 (2H, m, 7-H_2), 2.23 (0.3H,

q, *J* 7.5, 14-H_2), 1.04 (2H, t, *J* 7.5, 15-H_3), 0.93 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 205.0 (C-13), 193.8 (C-11), 176.0 (C-5), 174.0 (C-8), 77.3 (C-4), 70.4 (C-1), 40.4 (C-2), 39.8 (C-9), 37.3 (C-14), 36.4 (C-7), 36.3 (C-6), 29.7 (C-10), 21.3 (C-3), 20.9 (C-3), 7.7 (C-15). *m/z* (ESI^+) calc. for $[C_{16}H_{28}N_2O_6SNa]$ 399.1560 found 399.1543.

_S_-(13-Methylbut-12-enoyl)-pantetheine 64



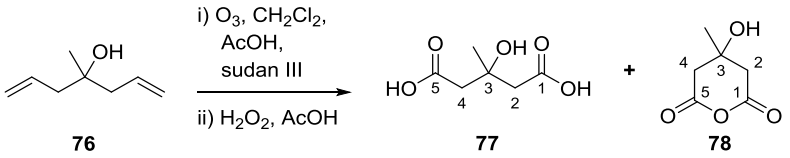
To thiol **59** (0.15 g, 0.47 mmol), acid 3,3-dimethylacrylic acid **68** (57 mg, 0.57 mmol) and EDCI (0.14 g, 0.71 mmol) under a nitrogen atmosphere was added CH_2Cl_2 (5 mL) and DIPEA (0.16 mL, 0.94 mmol). The reaction mixture was stirred at room temperature for 16 h then concentrated under reduced pressure. The crude oil was purified by column chromatography (EtOAc) furnishing a colourless oil (95 mg, 51%). To the protected pantetheine (80 mg, 0.20 mmol) was added 2:1 AcOH: H_2O (8 mL), the reaction mixture stirred at room temperature for 5 h then concentrated under reduced pressure. The crude oil was purified by column chromatography (5-7% MeOH:CH_2Cl_2) furnishing **64** as a colourless oil (66 mg, 92%). $[\alpha]_D^{24}$ +24.0 (*c* 1.0, MeOH). δ_H (400 MHz, CD_3OD) 6.04 (1H, s, 12-H), 3.88 (1H, s, 4-H), 3.52-3.44 (2H, m, 6-H_2), 3.44-3.39 (2H, m, 1-H_2), 3.37-3.31 (2H, m, 9-H_2), 3.01 (2H, t, *J* 7.0, 10-H_2), 2.40 (2H, t, *J* 7.0, 7-H_2), 2.14 (3H, s, 14-H_3), 1.89 (3H, s, 15-H_3), 0.91 (6H, s, $(CH_3)_2$). δ_C (100 MHz, CD_3OD) 190.0 (C-11), 176.0 (C-5), 173.9 (C-8), 155.8 (C-13), 123.9 (C-12), 77.3 (C-4), 70.4 (C-1), 40.4 (C-2), 40.3 (C-9), 36.4 (C-7), 36.3 (C-6), 29.0 (C-10), 27.2 (C-15), 21.3 (C-3), 21.2 (C-14), 20.9 (C-3). ν_{max} (neat) / cm^{-1} 3296 (br, w), 2930 (w), 1629 (s), 1527 (s). *m/z* (ESI^+) calc. for $[C_{16}H_{28}N_2O_5SNa]$ 383.1617 found 383.1623.

4-Methylhepta-1,6-dien-4-ol 76

To freshly distilled ethyl acetate **75** (1.80 g, 20.5 mmol) in dry THF (10 mL) at 0 °C under a nitrogen atmosphere was added allylmagnesium bromide (45 mL, 45 mmol, 1.0 M in

Et_2O). The reaction mixture was allowed to warm to room temperature, stirred for 16 h then quenched with 1 M HCl (30 mL). The aqueous phase was extracted with Et_2O (3 x 30 mL), the combined organic phases were then dried over $MgSO_4$ and concentrated *in vacuo*. The crude yellow oil was purified by column chromatography (10% EtOAc:petrol) furnishing **76** as a colourless oil (1.05 g, 41%). δ_H (400 MHz, $CDCl_3$) 5.87 (2H, ddt, *J* 17.5, 10.5, 7.5, 2-H and 6-H), 5.17-5.07 (4H, m, 1-H_2 and 7-H_2), 2.22 (4H, m, 3-H_2 and 5-H_2), 1.17 (3H, s, CH_3). δ_C (100 MHz, $CDCl_3$) 134.0 (C-2 and C-6), 118.9 (C-1 and C-7), 71.8 (C-4), 46.3 (C-3 and C-5), 26.8 (CH_3). *m/z* (ESI^+) calc. for [$C_8H_{14}O$] 149.1 found 149.1. The data are in accordance with the literature.[188]

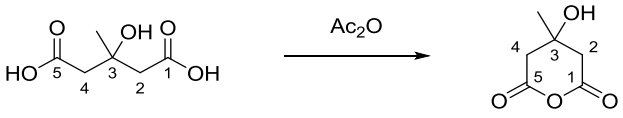
3-Hydroxy-3-methylpentanedioic acid 77 and HMG-anhydride 78



Alcohol **76** (0.50 g, 3.96 mmol) was dissolved in CH_2Cl_2 (10 mL) and AcOH (1 mL) at -78 °C and the tip of a spatula of sudan-III dye was added. Ozone was passed through the solution until the red colour dissipated then the reaction mixture was allowed to warm to room temperature. The solvent was removed under a stream of nitrogen. To the crude oil was added AcOH (20 mL) and 30% solution of H_2O_2 (13 mL), the reaction mixture was heated to reflux and stirred for 16 h. After cooling to 0 °C, the reaction was quenched with sat. aq. Na_2SO_3 (20 mL), extracted with EtOAc (2 x 20 mL), the combined organic extracts washed with brine (15 mL), dried over $MgSO_4$ and concentrated *in vacuo*. Diacid **77** was isolated as a white solid that didn't require any further purification (97 mg, 15%). δ_H (400 MHz, $(CD_3)_2CO$) 2.69 (4H, s, 2-H_2 and 4-H_2), 1.38 (3H, s, CH_3). δ_C (100 MHz, $(CD_3)_2CO$) 173.2 (CO), 70.1 (C-3), 45.2 (C-2 and C-4), 27.6 (CH_3). *m/z* (ESI^+) calc. for [$C_6H_{10}O_5Na$] 185.0420 found 185.0414.

The combined aqueous extracts were concentrated *in vacuo*, and to the resulting white solid at 0 °C was added Ac_2O (20 mL). The reaction mixture was allowed to warm to room temperature and stirred for 16 h. The subsequent red reaction mixture was concentrated under reduced pressure, the residue dissolved in CH_2Cl_2 and washed with brine (10 mL), dried over $MgSO_4$ and concentrated *in vacuo* furnishing **78** (102 mg, 15%) as a red solid. δ_H (400 MHz, $(CD_3)_2CO$) 4.65 (1H, br s, OH), 3.01 (2H, d, *J* 16.0, 2-*H*H and 4-*H*H), 2.86 (2H, d, *J* 16.0, 2-H*H* and 4-H*H*), 1.43 (3H, s, CH_3). δ_C (100 MHz,

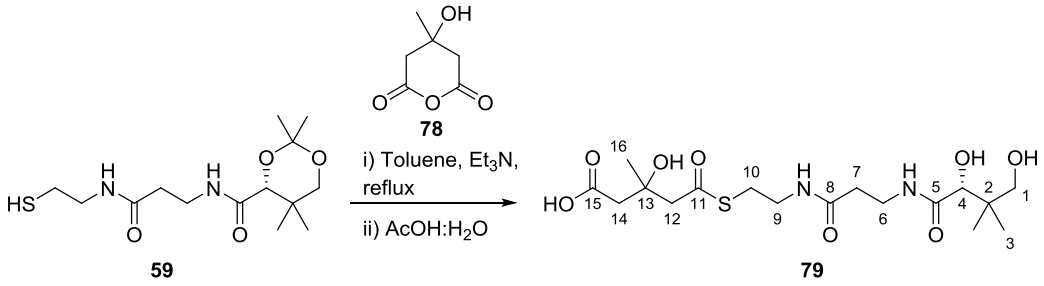
$(CD_3)_2CO$) 167.4 (C-1 and C-5), 67.8 (C-3), 44.3 (C-2 and C-4), 27.9 (CH_3). *m/z* (ESI^+) calc. for [$C_6H_8O_4Na$] 167.0315 found 167.0310. The data are in accordance with the literature.[188]



77 **78**

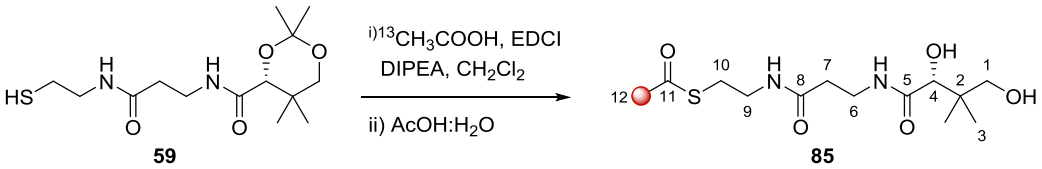
Diacid **77** (55 mg, 0.34 mmol) and Ac_2O (2 mL) were stirred at room temperature for 16 h. The reaction mixture was concentrated under reduced pressure furnishing **78** as a grey solid (48 mg, 99%). Spectroscopic data were in accordance with the data above.

S-(3-Hydroxy-3-methylglutaryl)-pantetheine 79



To thiol **59** (30 mg, 92.5 µmol) and anhydride **78** (20 mg, 139 µmol) under a nitrogen atmosphere was added dry toluene (6 mL) and Et_3N (5 drops). The reaction mixture was stirred at room temperature for 1 h then the solvent removed under reduced pressure. AcOH (4 mL) and H_2O (2 mL) were added and the reaction mixture was stirred for 5 h, then the solvent removed under reduced pressure. The crude red oil was purified by column chromatography (10% MeOH:CH_2Cl_2 + 0.1% AcOH) furnishing **79** as a colourless oil (21 mg, 54%). δ_H (400 MHz, CD_3OD) 3.88 (1H, s, 4-H), 3.53-3.46 (2H, m, 6-H_2), 3.46-3.40 (2H, m, 1-H_2), 3.38-3.33 (2H, m, 9-H_2), 3.01 (2H, t, *J* 6.5, 10-H_2), 2.95 (2H, d, *J* 3.0, 14-H_2), 2.61 (2H, s, 12-H_2), 2.40 (2H, t, *J* 6.5, 7-H_2), 1.35 (3H, s, 16-H_3), 0.91 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 198.3 (C-11), 176.1 (C-5), 175.2 (C-15), 173.9 (C-8), 77.3 (C-4), 71.1 (C-13), 70.4 (C-1), 55.0 (C-14), 46.1 (C-12), 40.4 (C-2), 39.9 (C-9), 36.4 (C-7), 36.3 (C-6), 29.5 (C-10), 27.8 (C-16), 21.3 (C-3). 20.9 (C-3). *m/z* (ESI^-) calc. for [$C_{17}H_{29}N_2O_8S$] 421.1650 found 421.1653.

S-[^{13}C]-Acetyl pantetheine 85



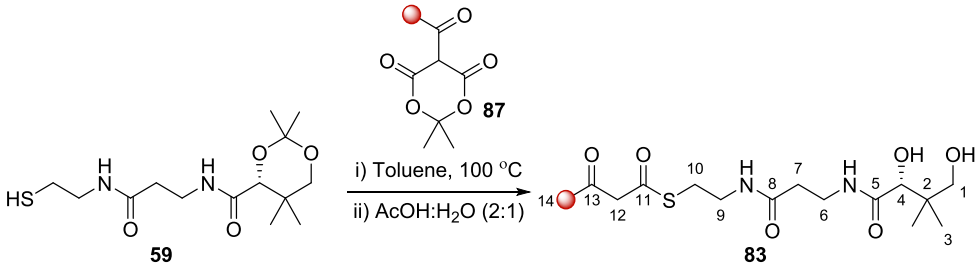
To thiol **59** (0.10 g, 0.34 mmol) in CH_2Cl_2 under a nitrogen atmosphere was added [2-^{13}C]-acetic acid (25 µL, 0.41 mmol), EDCI (90 mg, 0.50 mmol) and DIPEA (120 µL, 0.68 mmol.). The reaction mixture was stirred for 16 h, then the solvent removed under reduced pressure. The crude colourless oil was purified by column chromatography (50-80% EtOAc:hexane) furnishing a colourless oil (75 mg, 60%). To the oil (50 mg, 0.14 mmol) was added acetic acid (6 mL) and water (3 mL), the solution was stirred at room temperature for 3 h then concentrated *in vacuo*. The crude oil was purified by column chromatography (80% EtOAc:hexane) furnishing **85** as a colourless oil (44 mg, 98%). δ_H (500 MHz, CD_3OD) 3.90 (1H, s, 4-H), 3.55-3.48 (2H, m, 6-H_2), 3.47-3.40 (2H, m, 1-H_2), 3.40-3.34 (2H, m, 9-H_2), 3.01 (2H, t, *J* 6.5, 10-H_2), 2.42 (2H, t, *J* 6.5, 7-H_2), 2.34 (3H, d, *J* 130 Hz, 12-H_3), 0.93 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 197.0 (d, *J* 48.0 Hz, C-11), 176.0 (C-5), 173.9 (C-8), 77.3 (C-4), 70.3 (C-1), 40.4 (C-2), 40.0 (C-9), 36.4 (C-7), 36.3 (C-6), 30.5 (enhanced C-12), 29.5 (C-10), 21.3 (C-3), 20.9 (C-3). *m/z* (ESI$^+$) calc. for [$^{13}C_1C_{12}H_{24}N_2O_5SNa$] 344.1332 found 344.1345.

S-[^{13}C]-Acetyl Meldrum's Acid 87

To Meldrum's acid **65** (0.20 g, 1.38 mmol) in dry CH_2Cl_2 (5 mL) at 0 °C under a nitrogen atmosphere was added 2-[^{13}C]-acetyl chloride (118 µL, 1.67 mmol) and DMAP (0.34 g, 2.78 mmol). The reaction mixture was allowed to warm to room temperature and stirred for 3 h, then quenched by the addition of 1 M HCl (15 mL). The aqueous phase was extracted with CH_2Cl_2 (2 x 15 mL), the combined organic phases washed with brine (10 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified by column chromatography (50% EtOAc:petrol) furnishing **87** as a yellow solid (203 mg, 79%). δ_H (500 MHz, $CDCl_3$) 15.13 (1H, s, OH), 2.66 (3H, d, *J* 131.0, 1-H_3), 1.74 (6H, s,

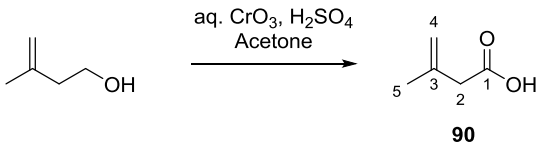
$(CH_3)_2$). δ_C (125 MHz, $CDCl_3$) 194.6 (C-2), 170.2 (CO), 160.5 (CO), 104.9 (C-6), 91.9 (C-3), 26.8 (C-7), 23.5 (enhanced C-1). *m/z* (ESI^+) calc. for [$^{13}C_1C_7H_{10}O_5Na$] 210.0454 found 210.0446.

_S_-[^{13}C]-Acetoacetyl-pantetheine 83



To thiol **59** (0.10 g, 0.31 mmol) and Meldrum's acid **87** (70 mg, 0.38 mmol) under a nitrogen atmosphere was added dry toluene (10 mL). The reaction mixture was heated to reflux and stirred for 16 h. Following cooling to room temperature the solvent was removed under reduced pressure and the crude oil was purified by column chromatography (4% MeOH:CH_2Cl_2) furnishing a colourless oil (65 mg, 52%). To the protected pantetheine (52 mg, 0.13 mmol) was added 2:1 AcOH:H_2O (4.5 mL) and the reaction mixture stirred at room temperature for 16 h. The solvent was removed under reduced pressure and the crude oil was purified by column chromatography (8% MeOH:CH_2Cl_2) furnishing **83** as a colourless oil (35 mg, 74%). δ_H (500 MHz, CD_3OD) 3.90 (1H, s, 4-H), 3.54-3.48 (2H, m, 6-H_2), 3.48-3.41 (2H, m, 1-H_2), 3.40-3.35 (2H, m, 9-H_2), 3.07 (2H, t, *J* 6.5, 10-H_2), 2.43 (2H, t, *J* 6.5, 7-H_2), 2.24 (2H, d, *J* 128 Hz, 14-H_3 keto), 1.94 (0.6H, d, *J* 128 Hz, 14-H_3 enol), 0.93 (6H, s, $(CH_3)_2$). δ_C (125 MHz, CD_3OD) 202.5 (d, *J* 43.0 Hz, C-13), 193.7 (C-11), 176.0 (C-5), 174.0 (C-8), 77.3 (C-4), 70.4 (C-1), 40.4 (C-2), 39.8 (C-9), 36.4 (C-7), 36.3 (C-6), 30.3 (enhanced C-14 keto), 29.7 (C-10), 21.3 (C-3), 20.9 (C-3), 20.8 (enhanced C-14 enol). *m/z* (ESI^+) calc. for [$^{13}C_1C_{14}H_{26}N_2O_6SNa$] 386.1437 found 386.1451.

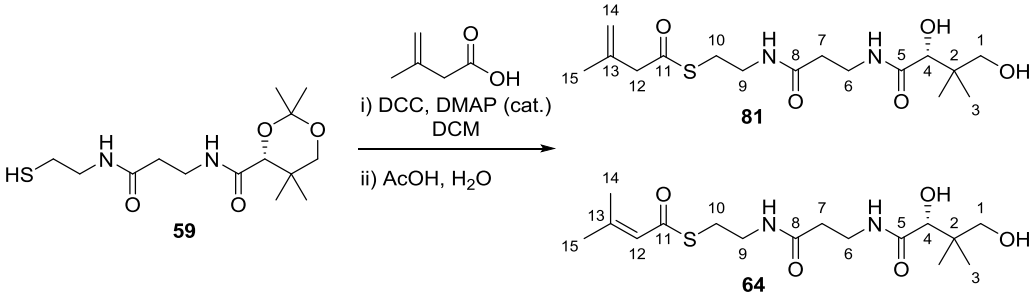
3-Methylbut-3-enoic acid 90



To 3-methylbut-3-en-1-ol (1.00 mL, 9.90 mmol) in acetone (50 mL) at 0 °C was added Jones reagent (5.20 mL, 2.67 M, 13.9 mmol) dropwise and the mixture stirred for 1 h. To

the solution was added Et_2O (30 mL) and 2 M NaOH (40 mL). The aqueous layer was then acidified to pH 2 with 1 M HCl and extracted with Et_2O (3 x 40 mL). The combined organic layers were washed with brine (2 x 20 mL), dried over $MgSO_4$ and the solvent removed *in vacuo* furnishing **90** as a pale-yellow oil (560 mg, 56%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.33 (1H, br s, COO*H*), 4.95 (1H, s, 4-*H*H), 4.88 (1H, s, 4-H*H*), 3.08 (2H, s, 2-H_2), 1.83 (3H, s, 5-H_3). ^{13}C NMR (100 MHz, $CDCl_3$) δ 177.6 (C-1), 138.1 (C-3), 115.5 (C-4), 43.2 (C-2), 22.5 (C-5). The data are in accordance with the literature.[189]

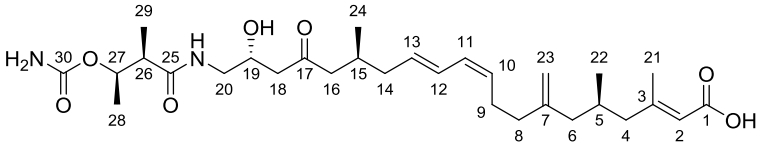
S-(13-Methylbut-13-enoyl) pantetheine 81 / S-(13-Methylbut-12-enoyl) pantetheine 64



To thiol **59** (50 mg, 0.16 mmol) in DCM (3 mL) at room temperature under a nitrogen atmosphere was added 3-methyl-3-butenoic acid **90** (16 µL, 17.7 mg, 0.16 mmol), N,N'-dicyclohexylcarbodiimide (29.2 µL, 39 mg, 0.19 mmol) and DMAP (4.8 mg, 39 µmol). The reaction mixture stirred for 2 h then the solution was filtered and washed with 1 M HCl (5 mL), sat. aq. $NaHCO_3$ (5 mL) and brine (10 mL). The combined organic phases were dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified by column chromatography (60% EtOAc:petrol) furnishing a colourless oil (40 mg, 57%) as a 1:1 mixture of *endo* and *exo* isomers. The mixture of isomers (5 mg, 12 µmol) was stirred in H_2O (2 mL) and AcOH (2 mL) at room temperature for 3 h. The solvent was removed *in vacuo* and the crude oil purified by column chromatography (70-100% EtOAc:petrol) furnishing **81** and **64** as a colourless oil (4 mg, 92%) as a 60:40 mixture of exo and endo isomers. ^{1}H NMR (600 MHz, 50 mM sodium phosphate, 100 mM NaCl, pH 8.0, 10% D_2O) δ 8.18 (1H, br s, NH), 8.05 (1H, br s, NH), 6.17 (1H, s, 12-H (endo)), 5.01 (1H, s, 14-*H*H (exo)), 3.98 (1H, s, 4-H (exo)), 3.97 (1H, s, 4-H (endo)), 3.52 – 3.44 (3H, obscured multiplet, 1-*H*H, 6-H_2), 3.41 – 3.36 (5H, obscured multiplet, 1-H*H*, 9-H_2 12-H_2 (exo)), 3.05 (2H, t, *J* 6.5, 10-H_2), 2.46 (overlapping t, *J* 6.5, 7-H_2 (endo and exo)), 2.11 (3H, s, 14-H_3 (endo)), 1.90 (3H, s, 15-H_3 (endo)), 1.75 (3H, s, 15-H_3 (exo)), 0.90 (3H, s, 3-H_3), 0.87 (3H, s, 3-H_3). 14-H*H* obscured by solvent suppression. ^{13}C NMR (150 MHz, D_2O)

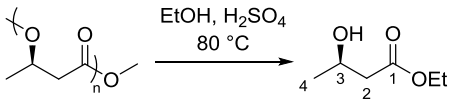
δ 204.7 (C-11 (exo)), 195.4 (C-11 (endo)), 177.8 (C-5), 176.7 (C-8), 160.2 (C-13 (endo)), 141.9 (C-13 (exo)), 124.9 (C-12 (endo)), 118.4 (C-14 (exo)), 78.5 (C-4), 71.2 (C-1), 54.7 (C-12 (exo)), 41.6 (C-9 (endo)), 41.3 (C-9 (exo)), 41.3 (C-2), 38.3 (C-6), 38.2 (C-7 (exo)), 38.0 (C-7 (endo)), 30.9 (C-10 (exo)), 30.6 (C-10 (endo)), 29.2 (C-15 (endo)), 24.2 (C-15 (exo)), 23.5 (C-14 (endo)), 23.2 (C-3), 21.8 (C-3); ν_{max} (thin film)/cm^{-1} 3290 (br), 2940 (m), 2800 (m), 1648 (s), 1536 (s), 1047 (s); *m/z* (ESI+) HRMS: Calculated for $C_{16}H_{29}N_2O_5S$ [M+H$^+$] 361.1792, found 361.1790.

Kalimantacin A



$[\alpha]_D^{24}$ -20.0 (*c* 1.0, MeOH). δ_H (600 MHz, $CDCl_3$) 6.28-6.25 (1H, m, NH), 6.24 (1H, m, 12-H), 5.95 (1H, t, *J* 11.0, 11-H), 5.70 (1H, s, 2-H), 5.59 (1H, dt, *J* 15.0, 7.0, 13-H), 5.31 (1H, dt, *J* 11.0, 7.5, 10-H), 4.90 (1H, dq, *J* 8.0, 6.0, 27-H), 4.79 (1H, s, 23-*H*H), 4.72 (1H, s, 23-H*H*), 4.19-4.15 (1H, m, 19-H), 3.44 (1H, ddd, *J* 14.0, 7.0 5.5, 20-*H*H), 3.31 (1H, ddd, *J* 14.0, 5.0, 3.0, 20-H*H*), 2.59 (1H, dd, *J* 18.0, 4.0, 18-*H*H), 2.54 (1H, dd, *J* 18.0, 9.0, 18-H*H*), 2.47 (1H, overlapping-m, 26-H), 2.43 (1H, overlapping-m, 16-*H*H), 2.30 (1H, overlapping-m, 9-*H*H), 2.27 (1H, overlapping-m, 16-H*H*), 2.23 (1H, overlapping-m, 9-H*H*), 2.15 (3H, d, *J* 1.0, 21-H_3), 2.10 (1H, overlapping-m, 6-*H*H), 2.09 (1H, overlapping-m, 15-H), 2.08 (1H, overlapping-m, 14-*H*H), 2.07 (2H, overlapping-m, 8-H_2), 2.06 (2H, overlapping-m, 4-H_2), 1.97-1.92 (1H, m, 14-H*H*), 1.91-1.86 (1H, m, 5-H), 1.68 (1H, dd, *J* 14.0, 9.5, 6-H*H*), 1.29 (3H, d, *J* 6.0, 28-H_3), 1.15 (3H, d, *J* 7.0, 29-H_3), 0.89 (3H, d, *J* 6.5, 24-H_3), 0.87 (3H, d, *J* 6.5, 22-H_3). δ_C (150 MHz, $CDCl_3$) 211.9 (C-17), 174.6 (C-25), 169.3 (C-1), 160.8 (C-3), 157.4 (C-30), 147.2 (C-7), 132.4 (C-13), 129.8 (C-10), 128.9 (C-11), 127.5 (C-12), 116.7 (C-2), 112.1 (C-23), 73.8 (C-27), 67.4 (C-19), 50.4 (C-16), 49.1 (C-4), 47.0 (C-26), 46.9 (C-18), 44.2 (C-20), 42.9 (C-6), 40.1 (C-14), 35.5 (C-8), 29.3 (C-15), 28.9 (C-5), 26.3 (C-9), 20.0 (C-22), 19.7 (C-24), 19.0 (C-21), 18.1 (C-28), 13.8 (C-29). ν_{max} (neat) / cm^{-1} 2957 (br, m), 2926 (br, m), 1731 (s), 1235 (m), 1164 (m). *m/z* (ESI$^+$) calc. for [$C_{30}H_{48}N_2O_7Na$] 571.3354 found 571.3350. The data are in accordance with the literature.[78]

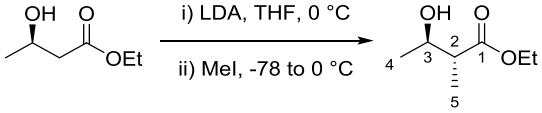
Ethyl (*R*)-3-hydroxybutanoate 97



104 **97**

Poly-(*R*)-3-hydroxybutanoate **104** (10.0 g), ethanol (130 mL) and concentrated sulfuric acid (4 mL) were heated to 80 °C and stirred for 7 days. The reaction mixture was cooled to room temperature and diluted with CH_2Cl_2 (60 mL), washed with brine (30 mL) and sat. aq. $NaHCO_3$ (30 mL). The combined aqueous phases were extracted with CH_2Cl_2 (3 x 70 mL), organic extracts were combined, dried over $MgSO_4$ and concentrated *in vacuo*. The crude oil was purified by vacuum distillation at 80 °C/12 mm Hg, furnishing **97** as a colourless oil (11.2 g, 73 %). $[\alpha]_D^{24}$ -43.5 (*c* 2.0, $CHCl_3$), [lit.[190] $[\alpha]_D^{24}$ -43.2 (*c* 1.0, $CHCl_3$)]. δ_H (400 MHz, $CDCl_3$) 4.21-4.18 (1H, m, 3-H), 4.15 (2H, q, *J* 7.0, $OC\mathit{H}_2CH_3$), 3.05-2.78 (1H, br s, OH), 2.47 (1H, dd, *J* 16.5, 3.5, 2-*H*H), 2.39 (1H, dd, *J* 16.5, 8.5, 2-H*H*), 1.26 (3H, t, *J* 7.0, $OCH_2C\mathit{H}_3$), 1.21 (3H, d, *J* 6.0, 4-H_3). δ_C (100 MHz, $CDCl_3$) 172.9 (C-1), 64.3 (C-3), 60.7 ($O\mathit{C}H_2CH_3$), 42.9 (C-2), 22.5 (C-4), 14.2 ($OCH_2\mathit{C}H_3$). ν_{max} (neat) / cm^{-1} 3425 (br), 2975 (w), 1731 (s), 1178 (s). The data are in accordance with the literature.[190]

Ethyl (2*R*, 3*R*)-3-hydroxy-2-methylbutanoate 105

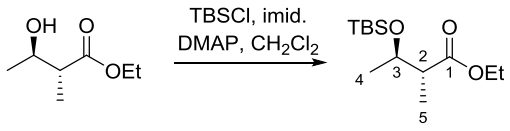


97 **105**

Butyllithium (83.2 mmol, 51.4 mL, 1.62 M in hexanes) was added dropwise to diisopropylamine (8.40 g, 83.2 mmol, 11.6 mL) in dry THF (50 mL) at 0 °C under a nitrogen atmosphere. The reaction mixture was stirred for 1 h, cooled to – 78 °C and alcohol **97** (5.00 g, 37.8 mmol) in dry THF (10 mL) was added dropwise. After stirring for 1.5 h, methyl iodide (6.40 g, 45.4 mmol, 2.82 mL) was added and the reaction mixture allowed to warm to 0 °C. After stirring for 4 h, the reaction was quenched with sat. aq. NH_4Cl (30 mL), acidified to pH 2 with 2 M HCl and extracted with EtOAc (3 x 60 mL). The combined organic extracts were washed with brine (40 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude brown oil was purified by column chromatography (10-33% EtOAc:petrol), furnishing **105** as a colourless oil (4.20 g, 76 %). $[\alpha]_D^{24}$ -27.0 (*c* 1.0,

$CHCl_3$), [lit.[191] $[\alpha]_D^{24}$ -30.3 (*c* 1.0, $CHCl_3$)]. δ_H (400 MHz, $CDCl_3$) 4.17 (2H, q, *J* 7.0, OCH_2CH_3), 3.88 (1H, quintet, *J* 6.5, 3-H), 2.44 (1H, quintet, *J* 7.0, 2-H), 1.28 (3H, t, *J* 7.0, OCH_2CH_3), 1.22 (3H, d, *J* 6.5, 4-H_3), 1.19 (3H, d, *J* 7.0, 5-H_3). δ_C (100 MHz, $CDCl_3$) 175.9 (C-1), 69.4 (C-3), 60.6 (OCH_2CH_3), 46.9 (C-2), 20.7 (C-4), 14.3 (OCH_2CH_3), 14.2 (C-5). ν_{max} (thin film) / cm^{-1} 3442 (br, w), 1714 (s), 1183 (s). *m/z* (ESI^+) calc. for [$C_7H_{14}O_3Na$] 169.1 found 169.1. The data are in accordance with the literature.[191]

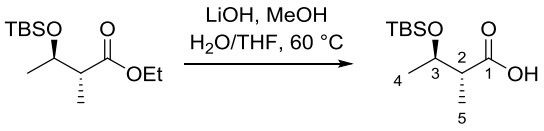
Ethyl (2*R*, 3*R*)-3-(*tert*-Butyldimethylsilyloxy)-2-methylbutanoate 106



105 **106**

To ester **105** (3.00 g, 20.5 mmol) in dry CH_2Cl_2 (40 mL) under a nitrogen atmosphere was added TBSCl (3.70 g, 24.6 mmol), imidazole (4.20 g, 61.5 mmol) and DMAP (125 mg, 1.03 mmol). The reaction mixture was stirred at room temperature for 16 h, quenched with sat. aq. NH_4Cl (40 mL) and extracted with CH_2Cl_2 (3 x 30 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*. Column chromatography (10% EtOAc:petrol) furnished **106** as a colourless oil (4.95 g, 93%). $[\alpha]_D^{24}$ -35.0 (*c* 2.0, $CHCl_3$), [lit.[192,193] $[\alpha]_D^{24}$ +29.4 (*c* 1.0, $CHCl_3$) and $[\alpha]_D^{24}$ +31.1 (*c* 1.0, $CHCl_3$) for enantiomer]. δ_H (400 MHz, $CDCl_3$) 4.11 (2H, q, *J* 7.0, OCH_2CH_3), 4.02 (1H, quintet, *J* 6.5, 3-H), 2.48 (1H, quintet, *J* 7.0, 2-H), 1.26 (3H, t, *J* 7.0, OCH_2CH_3), 1.13 (3H, d, *J* 6.5, 4-H_3), 1.08 (3H, d, *J* 7.0, 5-H_3), 0.86 (9H, s, $SiC(CH_3)_3$), 0.06 (3H, s, $SiCH_3$), 0.03 (3H, s, $SiCH_3$). δ_C (100 MHz, $CDCl_3$) 175.1 (C-1), 70.1 (C-3), 60.1 (OCH_2CH_3), 48.1 (C-2), 25.7 ($SiC(CH_3)_3$), 20.5 (C-4), 17.9 ($SiC(CH_3)_3$), 14.2 (OCH_2CH_3), 12.6 (C-5), -4.3 ($SiCH_3$), -5.1 ($SiCH_3$). ν_{max} (neat) / cm^{-1} 2930 (br, w), 1736 (s). *m/z* (ESI^+) calc. for [$C_{13}H_{28}O_3SiNa$] 283.2 found 283.2. The data are in accordance with the literature.[192,193]

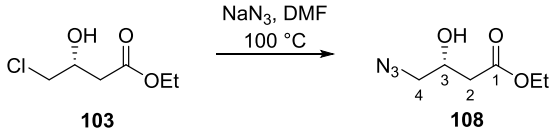
(2*R*, 3*R*)-3-(*tert*-Butyldimethylsilyloxy)-2-methylbutanoic acid 107



106 **107**

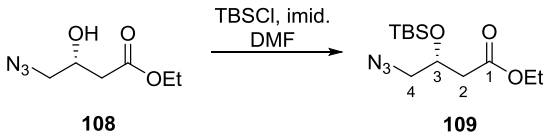
To ester **106** (0.20 g, 0.76 mmol) in THF (1 mL) and water (2 mL) was added anhydrous LiOH (55 mg, 0.23 mmol) in water (0.5 mL). After stirring at 60 °C for 16 h, LiOH monohydrate (0.16 g, 3.80 mmol) in water (1 mL) and MeOH (2 mL) were added and the reaction mixture stirred for 6 h. After cooling to room temperature, water (10 mL) was added and the mixture was extracted with Et_2O (2 x 10 mL). The aqueous phase was acidified to pH 3 with 6 M HCl and extracted with EtOAc (3 x 10 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*. The crude colourless oil was purified by column chromatography (5-10% EtOAc:petrol + 1% acetic acid), furnishing **107** as a colourless oil (120 mg, 68%). $[\alpha]_D^{24}$ -13.0 (*c* 1.0, $CHCl_3$), [lit.[194] $[\alpha]_D^{24}$ -12.5 (*c* 1.0, $CHCl_3$)]. δ_H (400 MHz, $CDCl_3$) 4.01 (1H, quintet, *J* 6.5, 3-H), 2.52 (1H, quintet, *J* 7.0, 2-H), 1.22 (3H, d, *J* 6.5, 4-H_3), 1.19 (3H, d, *J* 7.0, 5-H_3), 0.90 (9H, s, SiC(CH_3)$_3$), 0.11 (3H, s, $SiCH_3$), 0.09 (3H, s, $SiCH_3$). δ_C ($CDCl_3$, 100 MHz) 179.0 (C-1), 70.4 (C-3), 47.6 (C-2), 25.7 (SiC(*C*H_3)$_3$), 21.1 (C-4), 17.9 (Si*C*(CH_3)$_3$), 13.7 (C-5), -4.4 ($SiCH_3$), -5.2 ($SiCH_3$). *m/z* (ESI^+) calc. for [$C_{11}H_{24}O_3SiNa$] 255.1 found 255.1. The data are in accordance with the literature.[194]

Ethyl (*R*)-4-azido-3-hydroxybutanoate 108



To ester **103** (1.00 g, 6.00 mmol) in DMF (16 mL) at room temperature under a nitrogen atmosphere was added NaN_3 (0.78 g, 12.0 mmol). The reaction mixture was heated to 100 °C and stirred for 16 h. After cooling to room temperature, EtOAc (40 mL) was added and the reaction mixture washed with water (2 x 30 mL). The organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*. The crude orange oil was purified by column chromatography (EtOAc) furnishing **108** as an orange oil (0.98 g, 95 %). +7.0 (*c* 4.0, MeOH) [lit.[195,196] +7.1 (*c* 4.1, MeOH), $[\alpha]_D^{24}$ +7.4 (*c* 4.05, MeOH)]. δ_H (400 MHz, $CDCl_3$) 4.23-4.15 (3H, m, 3-H and OC$H_2$$CH_3$), 3.38-3.28 (2H, m, 4-$H_2$), 2.60-2.48 (2H, m, 2-$H_2$), 1.28 (3H, t, *J* 7.0, OCH_2CH_3). δ_C (100 MHz, $CDCl_3$) 172.1 (C-1), 67.4 (C-3), 61.0 (O*C*$H_2$$CH_3$), 55.5 (C-4), 38.4 (C-2), 14.1 (OCH_2*C*H_3). ν_{max} (thin film) / cm^{-1} 3451 (br), 2098 (s), 1717 (s), 1271 (m), 1174 (m). *m/z* (ESI^+) calc. for [$C_6H_{11}N_3O_3Na$] 196.1 found 196.1. The data are in accordance with the literature.[195,196]

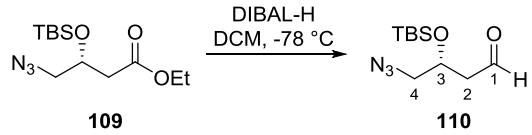
Ethyl (*R*)-4-Azido-3-(*tert*-butyldimethylsilyloxy) butanoate 109



108 **109**

To alcohol **108** (0.97 g, 5.60 mmol) in dry DMF (30 mL) at room temperature under a nitrogen atmosphere was added TBSCl (1.69 g, 11.2 mmol) and imidazole (0.87 g, 12.9 mmol) and stirred for 16 h. EtOAc (30 mL) was added and the reaction mixture washed with water (2 x 20 mL), brine (20 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude yellow oil was purified by column chromatography (10% EtOAc:petrol) furnishing **109** as a colourless oil (1.75 g, 99%). $[\alpha]_D^{24}$ +2.0 (*c* 1.0, $CHCl_3$) [lit.[195] $[\alpha]_D^{24}$ +2.2 (*c* 0.33, $CHCl_3$)]. δ_H (400 MHz, $CDCl_3$) 4.25 (1H, quintet, *J* 5.5, 3-H), 4.18-4.09 (2H, m, OCH_2CH_3), 3.38 (1H, dd, *J* 12.5, 4.0, 4-*H*H), 3.24 (1H, dd, *J* 12.5, 5.0, 4-H*H*), 2.60-2.50 (2H, m, 2-H_2), 1.27 (3H, t, *J* 7.0, OCH_2CH_3), 0.89 (9H, s, $SiC(CH_3)_3$), 0.13 (3H, s, $SiCH_3$), 0.09 (3H, s, $SiCH_3$). δ_C (100 MHz, $CDCl_3$) 170.8 (C-1), 68.7 (C-3), 60.6 (OCH_2CH_3), 56.4 (C-4), 40.1 (C-2), 25.6 ($SiC(CH_3)_3$), 17.9 ($SiC(CH_3)_3$), 14.1 (OCH_2CH_3), -4.8 ($SiCH_3$), -5.1 ($SiCH_3$). ν_{max} (thin film) / cm^{-1} 2930 (br, w), 2101 (s), 1734 (s), 836 (s). *m/z* (ESI^+) calc. for [$C_{12}H_{25}N_3O_3SiNa$] 310.2 found 310.2. The data are in accordance with the literature.[195]

Ethyl (*R*)-4-Azido-3-(*tert*-butyldimethylsilyloxy) butanal 110

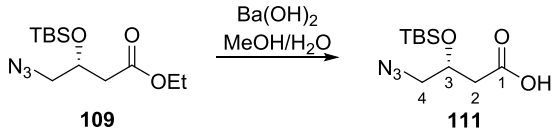


109 **110**

To ester **109** (0.20 g, 0.70 mmol) in dry CH_2Cl_2 (10 mL) at -78 °C under a nitrogen atmosphere was added DIBAL-H (0.90 mL, 0.90 mmol, 1.0 M in hexanes). The reaction mixture was stirred for 10 min then quenched with sat. aq. Rochelle's salt (10 mL) and stirred vigorously for 16 h. The mixture was extracted with CH_2Cl_2 (3 x 15 mL), the combined organic extracts dried over $MgSO_4$ and concentrated *in vacuo*. The crude yellow oil was purified by column chromatography (10% EtOAc:petrol) furnishing **110** as a colourless oil (0.16 g, 96%). $[\alpha]_D^{24}$ -6.0 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 9.81 (1H, s, 1-H), 4.34 (1H, quintet, *J* 5.0, 3-H), 3.39 (1H, dd, *J* 12.5, 5.0, 4-*H*H), 3.23 (1H, dd, *J* 12.5, 5.0, 4-H*H*), 2.62-2.74 (2H, m, 2-H_2), 0.90 (9H, s, $SiC(CH_3)_3$), 0.14 (3H, s, $SiCH_3$), 0.10 (3H, s, $SiCH_3$). δ_C (100 MHz, $CDCl_3$) 200.2 (C-1), 67.2 (C-3), 56.5 (C-4),

48.7 (C-2), 25.7 (SiC(CH_3)$_3$), 17.9 (SiC(CH_3)$_3$), -4.7 ($SiCH_3$), -5.0 ($SiCH_3$). *m/z* (ESI^+) calc. for [$C_{10}H_{21}N_3O_2SiNa$] 266.1 found 266.1. The compound has been reported in the literature but not fully characterised.[195]

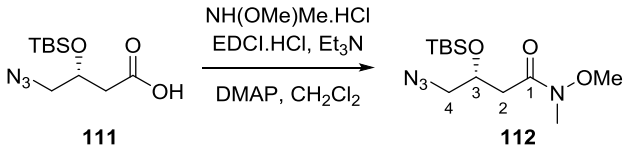
(*R*)-4-Azido-3-((*tert*-butyldimethylsilyl)oxy) butanoic acid 111



109 **111**

To ester **109** (0.50 g, 1.74 mmol) in MeOH (27 mL) and water (3 mL) was added $Ba(OH)_2$ (0.90 g, 5.22 mmol) and the reaction mixture was stirred at room temperature for 16 h. The MeOH was removed under reduced pressure and the residue diluted with water (5 mL) and EtOAc (15 mL). The phases were separated and the aqueous phase adjusted to pH 2 with 1 M HCl, extracted with EtOAc (4 x 15 mL) and the combined organic phases were washed with brine (15 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The light-yellow oil was purified by column chromatography (20% EtOAc:petrol) furnishing **111** as a colourless oil (366 mg, 82%). $[\alpha]_D^{24}$ +4.0 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 4.25 (1H, quintet, *J* 5.0, 3-H), 3.38 (1H, dd, *J* 12.5, 5.0, 4-*H*H), 3.26 (1H, dd, *J* 12.5, 5.0, 4-H*H*), 2.63 (1H, dd, *J* 16.0, 6.0, 2-*H*H), 2.57 (1H, dd, *J* 16.0, 6.5, 2-H*H*), 0.89 (9H, s, SiC(CH_3)$_3$), 0.13 (3H, s, $SiCH_3$), 0.10 (3H, s, $SiCH_3$). δ_C (100 MHz, $CDCl_3$) 176.6 (C-1), 68.4 (C-3), 56.2 (C-4), 39.8 (C-2), 25.6 (SiC(CH$_3$)$_3$), 17.9 (SiC(CH$_3$)$_3$), -4.7 ($SiCH_3$), -5.1 ($SiCH_3$). ν_{max} (neat) / cm^{-1} 2929 (m), 2857 (m), 2102 (s), 1711 (s), 1255 (m). *m/z* (ESI^+) calc. for [$C_{10}H_{21}N_3O_3SiNa$] 282.1244 found 282.1236. The compound is reported in the literature but not fully characterised.[197]

(*R*)-4-Azido-3-(*tert*-butyldimethylsilyloxy)-*N*-methoxy-*N*-methylbutanamide 112

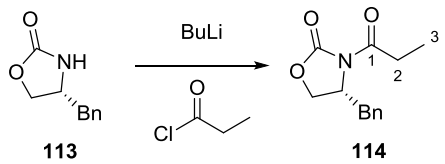


111 **112**

To acid **111** (0.35 g, 1.35 mmol) in CH_2Cl_2 (15 mL) under a nitrogen atmosphere was added *N*,*O*-dimethylhydroxylamine hydrochloride (0.66 g, 6.75 mmol) and EDCI.HCl (0.52 g, 2.70 mmol). DMAP (83 mg, 0.68 mmol) and triethylamine (0.94 mL, 6.75 mmol) were added and the reaction mixture stirred at room temperature for 4 h. The reaction was quenched with sat. aq. NH_4Cl (25 mL), the phases were separated and the aqueous phase

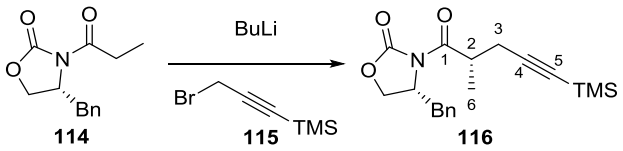
extracted with CH_2Cl_2 (2 x 20 mL). The combined organic phases were dried over $MgSO_4$ and concentrated *in vacuo*. The colourless residue was purified by column chromatography (20% EtOAc:petrol) furnishing **112** as a colourless oil (364 mg, 89%). $[\alpha]_D^{24}$ -25.0 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 4.40-4.33 (1H, m, 3-H), 3.70 (3H, s, OCH_3), 3.42 (1H, dd, *J* 12.5, 4.0, 4-*H*H), 3.19 (1H, overlapping dd, 4-H*H*), 3.16 (3H, s, NCH_3), 2.74 (1H, *J* 16.0, 6.0, 2-*H*H), 2.60 (1H, dd, *J* 16.0, 7.0, 2-H*H*), 0.89 (9H, s, $SiC(CH_3)_3$), 0.12 (3H, s, $SiCH_3$), 0.08 (3H, s, $SiCH_3$). δ_C (100 MHz, $CDCl_3$) 171.3 (C-1), 68.8 (C-3), 61.3 (OCH_3), 56.6 (C-4), 36.9 (C-2), 31.9 (NCH_3), 25.7 ($SiC(CH_3)_3$), 17.9 ($SiC(CH_3)_3$), -4.8 ($SiCH_3$), -4.9 ($SiCH_3$). ν_{max} (neat) / cm^{-1} 2929 (w), 2100 (s), 1660 (s), 1097 (s). *m/z* (ESI^-) calc. for $[C_{12}H_{26}N_4O_3SiNa]$ 325.1666 found 325.1658.

(*R*)-4'-Benzyl-*N*-propionyloxazolidin-2'-one 114



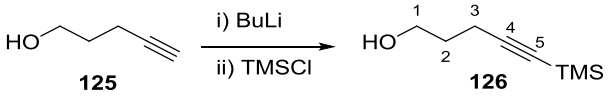
n-Butyllithium (4.50 mL, 6.77 mmol, 1.5 M in hexanes) was added dropwise to auxiliary **113** (1.00 g, 5.64 mmol) in dry THF (25 mL) at -78 °C under a nitrogen atmosphere. Propionyl chloride (0.54 mL, 6.20 mmol) was added and the reaction mixture stirred at -78 °C for 30 min, quenched with sat. aq. NH_4Cl (15 mL) and extracted with EtOAc (3 x 30 mL). The combined organic extracts were washed with sat. aq. $NaHCO_3$ (15 mL), brine (15 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude colourless oil was purified by column chromatography (10-20 % EtOAc:petrol) furnishing **114** as a waxy white solid (0.77 g, 58%). $[\alpha]_D^{24}$ -54.0 (*c* 1.0, $CHCl_3$), [lit.[198] $[\alpha]_D^{24}$ -65.1 (*c* 1.0, $CHCl_3$)]. δ_H (400 MHz, $CDCl_3$) 7.38-7.31 (2H, m, Ar-H), 7.31-7.27 (1H, m, Ar-H), 7.24-7.20 (2H, m, Ar-H), 4.68 (1H, ddt, *J* 10.5, 7.0, 3.0, NCH), 4.24-4.16 (2H, m, OCH_2), 3.32 (1H, dd, *J* 13.5, 3.0, C*H*HPh), 3.05-2.88 (2H, m, 2-H_2), 2.78 (1H, dd, *J* 13.5, 10.0, CH*H*Ph), 1.21 (3H, t, *J* 7.5, 3-H_3). δ_C (100 MHz, $CDCl_3$) 174.1 (C-1), 153.5 (OC(O)N), 135.3 (Ar-C), 129.4 (Ar-CH), 129.0 (Ar-CH), 127.3 (Ar-CH), 66.2 (OCH_2), 55.2 (NCH), 37.9 (CH_2Ph), 29.2 (C-2), 8.3 (C-3). ν_{max} (thin film) / cm^{-1} 1774 (s), 1698 (s), 1211 (m), 703 (m). *m/z* (ESI^+) calc. for $[C_{13}H_{15}NO_3Na]$ 256.1 found 256.1. The data are in accordance with the literature.[198]

(*R*)-4'-Benzyl-*N*-((*S*)-2-methyl-5-(trimethylsilyl)pent-4-ynoyl)oxazolidin-2'-one 116



To NaHMDS (0.47 mmol, 0.47 mL, 1.0 M in THF) in dry THF (4 mL) at -78 °C under a nitrogen atmosphere was added auxiliary **114** (0.10 g, 0.43 mmol) in dry THF (2 mL). The reaction mixture was stirred for 1 h, then bromide **115** (0.16 g, 0.86 mmol) in dry THF (2 mL) was added dropwise and stirred at -78 °C for 4 h. The reaction mixture was allowed to warm to room temperature over 16 h, quenched with sat. aq. NH_4Cl (15 mL) and extracted with CH_2Cl_2 (3 x 15 mL). The combined organic extracts were dried over $MgSO_4$ and concentrated *in vacuo*. The crude orange oil was purified by column chromatography (20% EtOAc:petrol) furnishing **116** as a waxy white solid (24 mg, 16%). δ_H (400 MHz, $CDCl_3$) 7.38-7.32 (2H, m, Ar-H), 7.31-7.28 (1H, m, Ar-H), 7.26-7.22 (2H, m, Ar-H), 4.71 (1H, ddt, *J* 10.5, 7.0, 3.0, NCH), 4.25-4.16 (2H, m, OCH_2), 3.95 (1H, sex, *J* 6.5, 2-H), 3.31 (1H, dd, *J* 13.5, 3.0, C*H*HPh), 2.77 (1H, dd, *J* 13.5, 9.5, CH*H*Ph), 2.63 (1H, dd, *J* 17.0, 7.0, 3-*H*H), 2.55 (1H, dd, *J* 17.0, 7.0, 3-H*H*), 1.28 (3H, d, *J* 7.0, 6-H_3), 0.13 (9H, s, $Si(CH_3)_3$). δ_C (100 MHz, $CDCl_3$,) 175.2 (C-1), 153.0 (OC(O)N), 135.2 (Ar-C), 129.4 (Ar-CH), 129.0 (Ar-CH), 127.4 (Ar-CH), 103.7 (C-5), 86.5 (C-4), 66.1 (OCH_2), 55.3 (NCH), 38.0 (*C*H_2Ph), 37.4 (C-2), 24.1 (C-3), 16.5 (C-6), 0.1 ($Si(CH_3)_3$).

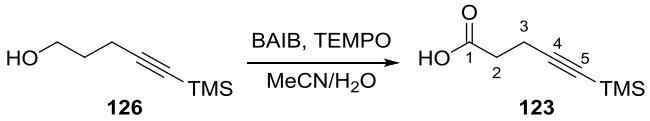
5-(Trimethylsilyl)pent-4-yn-1-ol 126



n-Butyllithium (34.4 mL, 49.9 mmol, 1.45 M in hexanes) was added to alcohol **125** (2.00 g, 23.8 mmol) in dry THF (40 mL) at -78 °C under a nitrogen atmosphere. The reaction mixture was stirred for 1 h, then TMSCl (5.40 g, 49.9 mmol, 6.30 mL) was added and allowed to warm to room temperature over 16 h. 1 M HCl (40 mL) was added, the reaction mixture stirred for 16 h then extracted with CH_2Cl_2 (2 x 30 mL). The combined organic extracts dried over $MgSO_4$ and concentrated *in vacuo*. The crude yellow oil was purified by column chromatography (20% EtOAc:petrol) furnishing **126** as a colourless oil (3.71 g, 99%). δ_H (400 MHz, $CDCl_3$,) 3.76 (2H, t, *J* 6.0, 1-H_2), 2.35 (2H, t, *J* 7.0, 3-H_2), 1.77 (2H, pent, *J* 6.5, 2-H_2), 0.15 (9H, s, $Si(CH_3)_3$). δ_C (100 MHz, $CDCl_3$) 106.6 (C-

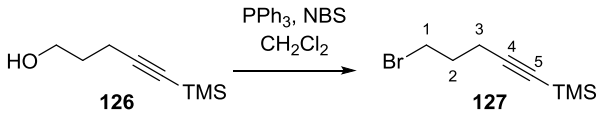
4), 85.3 (C-5), 61.9 (C-1), 31.1 (C-2), 16.6 (C-3), 0.1 ($Si(CH_3)_3$). ν_{max} (thin film) / cm^{-1} 3324 (br), 2956 (br, w), 2174 (w), 1248 (m), 838 (s). *m/z* (ESI^+) calc. for [$C_8H_{16}OSiNa$] 179.1 found 179.1. The data are in accordance with the literature.[199]

5-(Trimethylsilyl)pent-4-ynoic acid 123



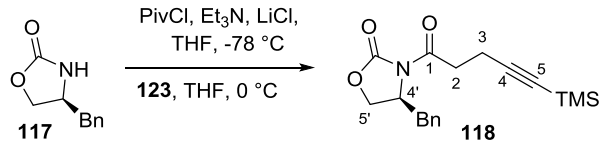
To alcohol **126** (1.80 g, 11.5 mmol) in MeCN (30 mL) and water (30 mL) at room temperature was added TEMPO (0.54 g, 3.46 mmol) and BAIB (9.30 g, 28.8 mmol) and stirred for 16 h. The reaction mixture was extracted with CH_2Cl_2 (30 mL), the aqueous phase acidified to pH 1 with 2 M HCl and extracted with CH_2Cl_2 (2 x 20 mL). The combined organic extracts were washed with brine (20 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude brown oil was purified by column chromatography (15% EtOAc:petrol + 0.5% acetic acid) furnishing **123** as an orange solid (1.34 g, 68%). δ_H (400 MHz, $CDCl_3$) 2.65-2.52 (4H, m, 2-H_2 and 3-H_2), 0.15 (9H, s, $Si(CH_3)_3$). δ_C (100 MHz, $CDCl_3$) 177.6 (C-1), 104.5 (C-4), 85.7 (C-5), 33.4 (C-2), 15.5 (C-3), 0.0 ($Si(CH_3)_3$). ν_{max} (thin film) / cm^{-1} 2958 (br, w), 2176 (w), 1710 (m). *m/z* (ESI^-) calc. for [$C_{16}H_{29}O_3Si_2$] 169.1 found 169.1. The data are in accordance with the literature.[200]

1-Bromo-5-(trimethylsilyl)pent-4-yne 127



To alcohol **126** (2.00 g, 12.8 mmol) in dry CH_2Cl_2 (50 mL) at 0 °C under a nitrogen atmosphere was added PPh_3 (4.00 g, 15.4 mmol) and NBS (2.70 g, 15.4 mmol). The reaction mixture was stirred for 3 h, Et_2O (50 mL) added and stirred for a further 16 h. The precipitate was filtered under vacuum and the filtrate concentrated *in vacuo*. Purification by column chromatography (10% EtOAc:petrol) furnished **127** as a colourless oil (2.80 g, 99%). δ_H (400 MHz, $CDCl_3$) 3.51 (2H, t, *J* 6.5, 1-H_2), 2.42 (2H, t, *J* 7.0 , 3-H_2), 2.05 (2H, pent, *J* 7.0, 2-H_2), 0.15 (9H, s, $Si(CH_3)_3$). δ_C (100 MHz, $CDCl_3$) 104.9 (C-4), 85.8 (C-5), 32.3 (C-1), 31.4 (C-2), 18.6 (C-3), 0.1 ($Si(CH_3)_3$). ν_{max} (thin film) / cm^{-1} 2959 (w), 2175 (w), 1739 (m), 1246 (m), 839 (s). The data are in accordance with the literature.[199]

(*S*)-4'-Benzyl-*N*-(5-(trimethylsilyl)pent-4-ynoyl)oxazolidin-2'-one 118



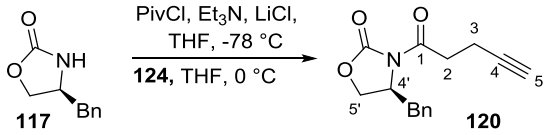
To 5-(trimethylsilyl)pent-4-ynoic acid **123** (0.24 g, 1.41 mmol) in dry THF (7 mL) at -78 °C under a nitrogen atmosphere was added Et_3N (0.49 mL, 3.52 mmol) and PivCl (0.17 mL, 1.41 mmol). The reaction mixture was allowed to warm to 0 °C over 1 h, then LiCl (89 mg, 2.11 mmol) and auxiliary **117** (0.25 g, 1.41 mmol) were added. After stirring at 0 °C for 1.5 h, the reaction mixture was quenched with sat. aq. NH_4Cl (15 mL) and extracted with EtOAc (3 x 10 mL). The combined organic extracts were washed with brine (10 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude orange oil was purified by column chromatography (30% EtOAc:petrol) furnishing **118** as a colourless oil (385 mg, 83%). $[\alpha]_D^{24}$ +75 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 7.37-7.31 (2H, m, Ar-H), 7.31-7.27 (1H, m, Ar-H), 7.23-7.19 (2H, m, Ar-H), 4.73-7.66 (1H, m, NCH), 4.26-4.17 (2H, m, OCH_2), 3.32-3.20 (2H, m, C*H*HPh and 2-*H*H), 3.18-3.09 (1H, m, 2-H*H*), 2.80 (1H, dd, *J* 13.5, 9.5, CH*H*Ph), 2.64 (2H, t, *J* 7.5, 3-H_2), 0.15 (9H, s, $Si(CH_3)_3$). δ_C (100 MHz, $CDCl_3$) 171.3 (C-1), 153.4 (NCO), 135.1 (Ar-C), 129.4 (Ar-CH), 129.0 (Ar-CH), 127.4 (Ar-CH), 105.0 (C-4), 85.4 (C-5), 66.3 (OCH_2), 55.1 (NCH), 37.8 (Ph*C*H_2), 34.9 (C-2), 15.0 (C-3), 0.0 ($Si(CH_3)_3$). ν_{max} (neat) / cm^{-1} 2970 (w, br), 2172 (w), 1788 (s), 1702 (s), 1193 (s). *m/z* (ESI^+) calc. for [$C_{18}H_{23}NO_3NaSi$] 352.1339 found 352.1349.

(*S*)-4'-Benzyl-*N*-((*S*)-2-methyl-5-(trimethylsilyl)pent-4-ynoyl)oxazolidin-2'-one 119

To auxiliary **118** (50 mg, 0.15 mmol) in dry THF (2 mL) at -78 °C under a nitrogen atmosphere was added NaHMDS (0.18 mmol, 0.18 mL, 1.0 M in THF). After stirring for 1 h, MeI (11 µL, 0.18 mmol) was added and stirring continued for a further 4 h. The reaction mixture was quenched with sat. aq. NH_4Cl (10 mL), extracted with CH_2Cl_2 (2 x 10 mL) and the combined organic extracts washed with brine (10 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude orange oil was purified by column chromatography (10-20% EtOAc:petrol) furnishing **119** as a white solid (20 mg, 40%). $[\alpha]_D^{24}$ +72.0 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 7.37-7.31 (2H, m, Ar-H), 7.31-7.27 (1H, m, Ar-H), 7.23-

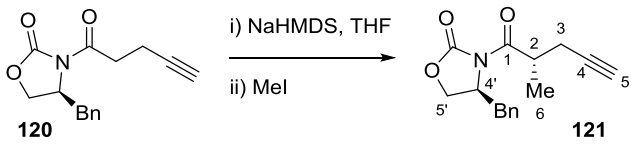
7.19 (2H, m, Ar-H), 4.72-4.66 (1H, m, NCH), 4.24-4.16 (2H, m, OCH_2), 3.92 (1H, sex, *J* 7.0, 2-H), 3.26 (1H, dd, *J* 13.5, 3.5, PhC*H*H), 2.79 (1H, dd, *J* 13.5, 9.5, PhCH*H*), 2.57 (1H, dd, *J* 17.0, 7.0, 3-*H*H), 2.46 (1H, dd, *J* 17.0, 7.5, 3-H*H*), 1.32 (3H, d, *J* 7.0, 6-H_3), 0.14 (9H, s, $Si(CH_3)_3$). δ_C (100 MHz, $CDCl_3$) 175.4 (C-1), 153.0 (NCO), 135.1 (Ar-C), 129.4 (Ar-CH), 128.9 (Ar-CH), 127.4 (Ar-CH), 104.0 (C-4), 86.3 (C-5), 66.2 (OCH_2), 55.3 (NCH), 37.9 (Ph*C*H_2), 37.5 (C-2), 23.8 (C-3), 17.0 (C-6), 0.0 ($Si(CH_3)_3$). ν_{max} (neat) / cm^{-1} 2959 (br,w), 2176 (w), 1781 (s), 1698 (m), 841 (s). *m/z* (ESI^+) calc. for [$C_{19}H_{25}NO_3NaSi$] 366.1496 found 366.1487.

(*S*)-4'-Benzyl-*N*-(pent-4-ynoyl)oxazolidin-2'-one 120



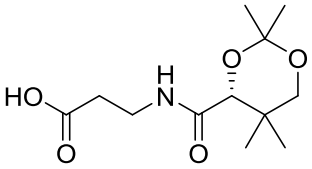
To pent-4-ynoic acid **124** (1.00 g, 10.2 mmol) in dry THF (50 mL) at -78 °C under a nitrogen atmosphere was added Et_3N (3.55 mL, 25.5 mmol) and PivCl (1.26 mL, 10.2 mmol). The reaction mixture was allowed to warm to 0 °C over 1 h, then LiCl (648 mg, 15.3 mmol) and auxiliary **117** (1.81 g, 10.2 mmol) were added. After stirring for 3 h, the reaction mixture was quenched with sat. aq. NH_4Cl (30 mL) and extracted with EtOAc (3 x 30 mL). The combined organic extracts were washed with brine (30 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude yellow oil was purified by column chromatography (20-40% EtOAc:petrol) furnishing **120** as a colourless oil (2.02 g, 77%). $[\alpha]_D^{24}$ +76.0 (*c* 1.0, $CHCl_3$), [lit.[201] $[\alpha]_D^{24}$ +98.6 (*c* 1.8, CH_2Cl_2)]. δ_H (400 MHz, $CDCl_3$) 7.37-7.31 (2H, m, Ar-H), 7.31-7.27 (1H, m, Ar-H), 7.23-7.19 (2H, m, Ar-H), 4.73-4.67 (1H, m, NCH), 4.26-4.17 (2H, m, OCH_2), 3.31 (1H, dd, *J* 13.5, 3.0, PhC*H*H), 3.27-3.13 (2H, m, 2-H_2), 2.80 (1H, dd, *J* 13.5, 9.5, PhCH*H*), 2.61 (2H, td, *J* 7.0, 3.0, 3-H_2), 2.01 (1H, t, *J* 3.0, 5-H). δ_C (100 MHz, $CDCl_3$) 171.2 (C-1), 153.4 (NCO), 135.1 (Ar-C), 129.4 (Ar-CH), 129.0 (Ar-CH), 127.4 (Ar-CH), 82.6 (C-4), 69.0 (C-5), 66.4 (OCH_2), 55.1 (NCH), 37.8 (Ph*C*H_2), 34.8 (C-2), 13.5 (C-3). ν_{max} (neat) / cm^{-1} 3290 (s), 2915 (br, w), 1766 (s), 1702 (s). *m/z* (ESI^+) calc. for [$C_{15}H_{15}NO_3Na$] 280.0944 found 280.0939. The data are in accordance with the literature.[201]

(*S*)-4'-Benzyl-*N*-((*S*)-2-methylpent-4-ynoyl)oxazolidin-2'-one 121

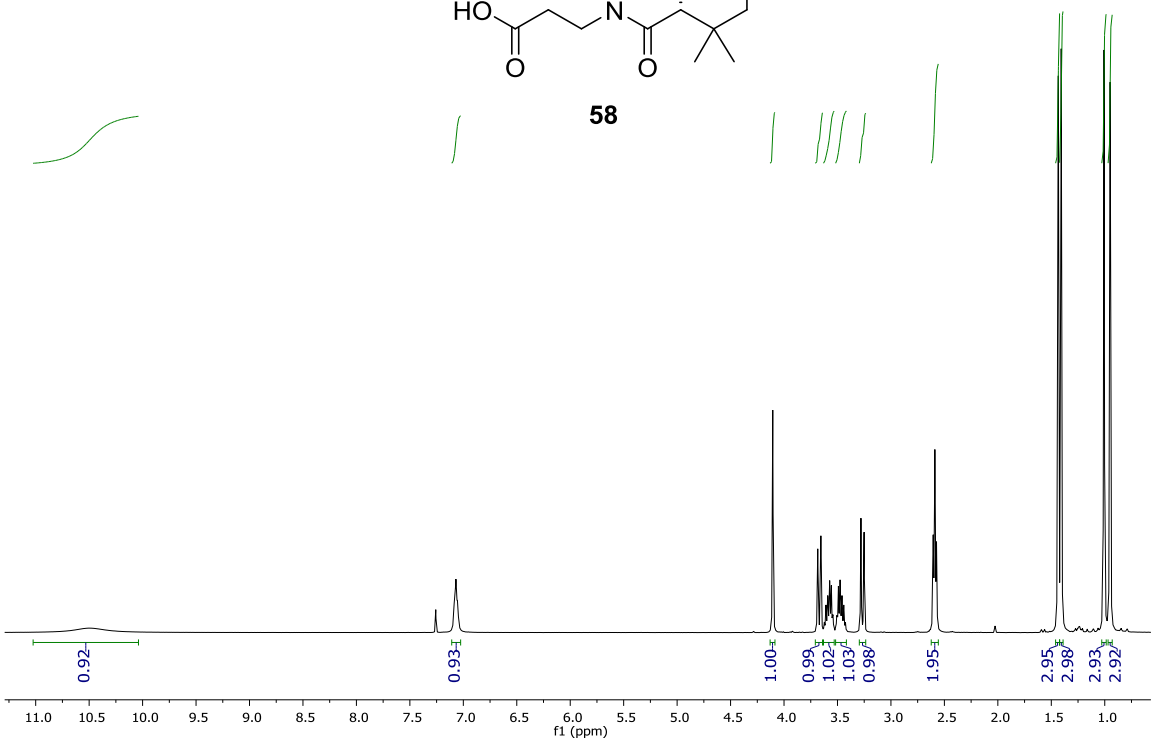


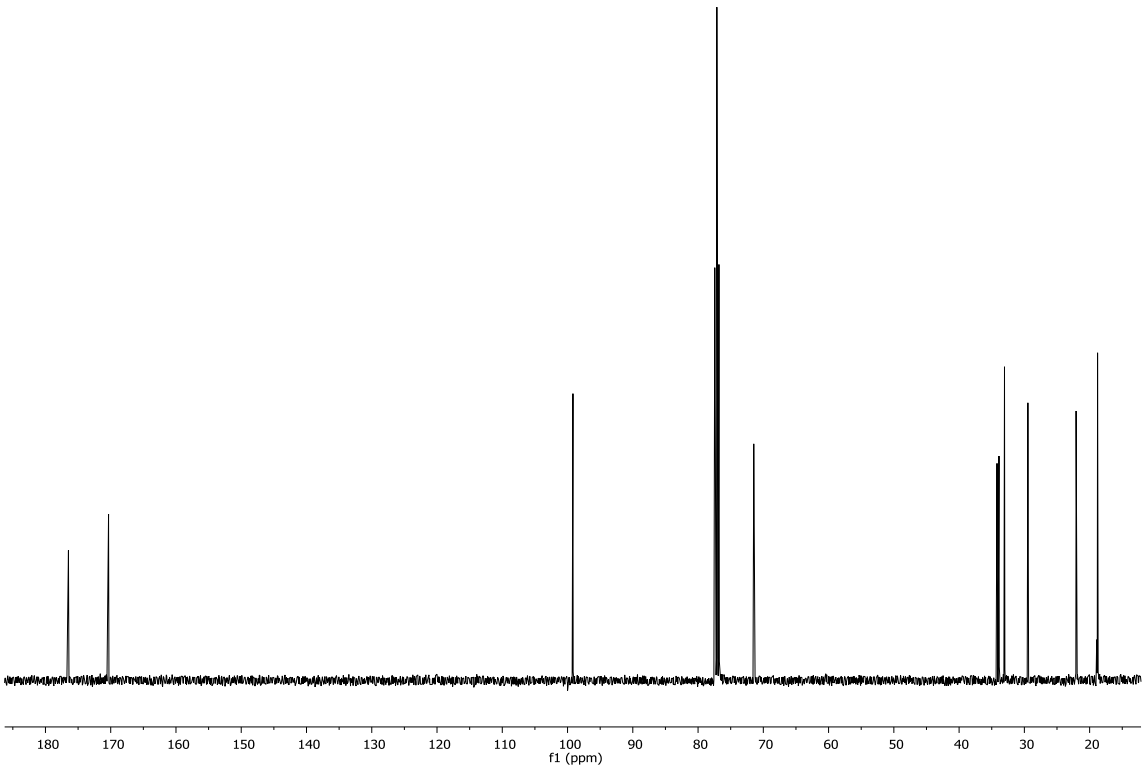
To NaHMDS (2.52 mmol, 2.52 mL, 1.0 M in hexanes) in dry THF (5 mL) at -78 °C under a nitrogen atmosphere was added auxiliary **120** (0.50 g, 1.94 mmol) in dry THF (5 mL) over 20 min. After stirring for 1 h, MeI (0.60 mL, 9.70 mmol) was added and stirring continued for a further 2 h at -50 °C. The reaction mixture was quenched with sat. aq. NH_4Cl (20 mL), extracted with CH_2Cl_2 (3 x 15 mL) and the combined organic extracts washed with brine (10 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude orange oil was purified by column chromatography (20% EtOAc:petrol) furnishing **121** as a white solid (273 mg, 52%). $[\alpha]_D^{24}$ +88 (*c* 1.0, $CHCl_3$). δ_H (400 MHz, $CDCl_3$) 7.37-7.31 (2H, m, Ar-H), 7.31-7.24 (1H, m, Ar-H), 7.23-7.19 (2H, m, Ar-H), 4.75-4.68 (1H, m, NCH), 4.26-4.17 (2H, m, OCH_2), 3.93 (1H, sex, *J* 7.0, 2-*H*), 3.27 (1H, dd, *J* 13.5, 3.5, PhC*H*H), 2.79 (1H, dd, *J* 13.5, 9.5, PhCH*H*), 2.58 (1H, ddd, *J* 17.0, 7.0, 3.0, 3-*H*H), 2.42 (1H, ddd, *J* 17.0, 7.0, 3.0, 3-H*H*), 1.99 (1H, t, *J* 3.0, 5-H), 1.33 (3H, t, *J* 7.0, 6-H_3). δ_C (100 MHz, $CDCl_3$) 175.2 (C-1), 153.0 (NCO), 135.1 (Ar-C), 129.4 (Ar-CH), 128.9 (Ar-CH), 127.4 (Ar-CH), 81.6 (C-4), 69.8 (C-5), 66.2 (OCH_2), 55.3 (NCH), 37.9 (Ph*C*H_2), 37.5 (C-2), 22.2 (C-3), 17.1 (C-6). ν_{max} (neat) / cm^{-1} 3264 (s), 1775 (s), 1695 (s), 1192 (s). *m/z* (ESI^+) calc. for [$C_{16}H_{17}NO_3Na$] 294.1101 found 294.1113.

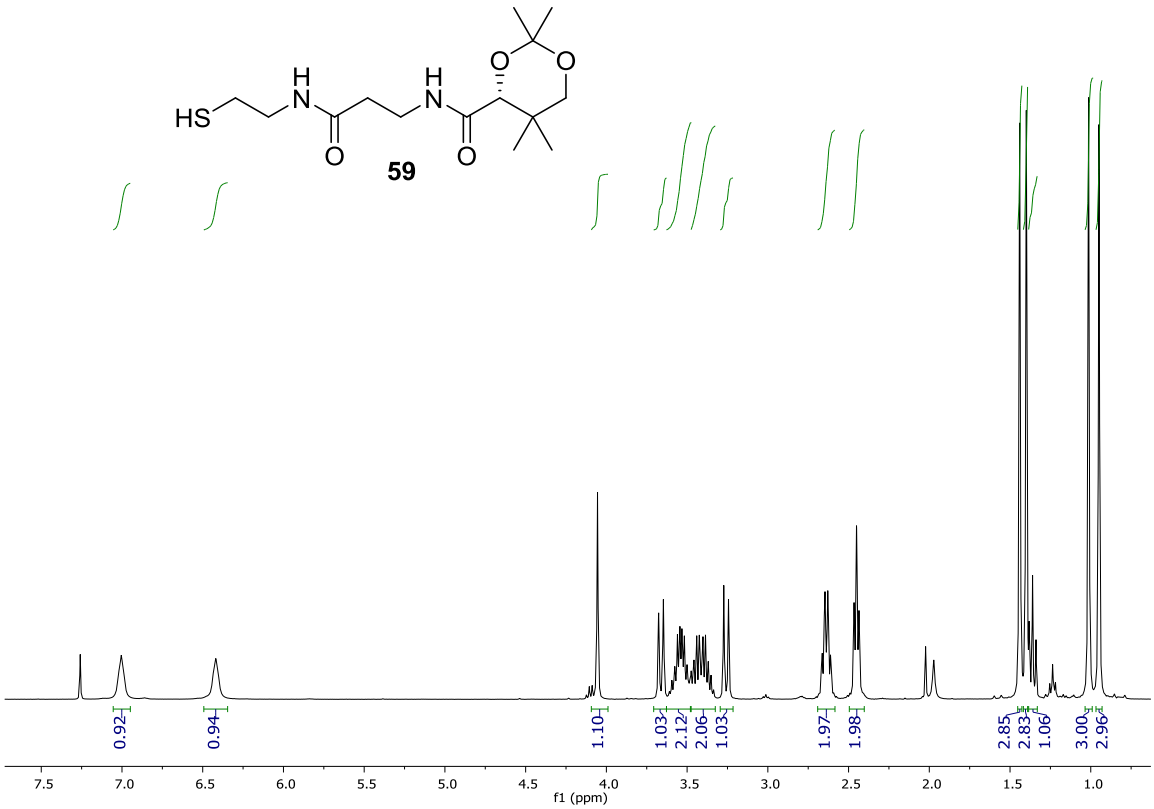
6 NMR Spectra

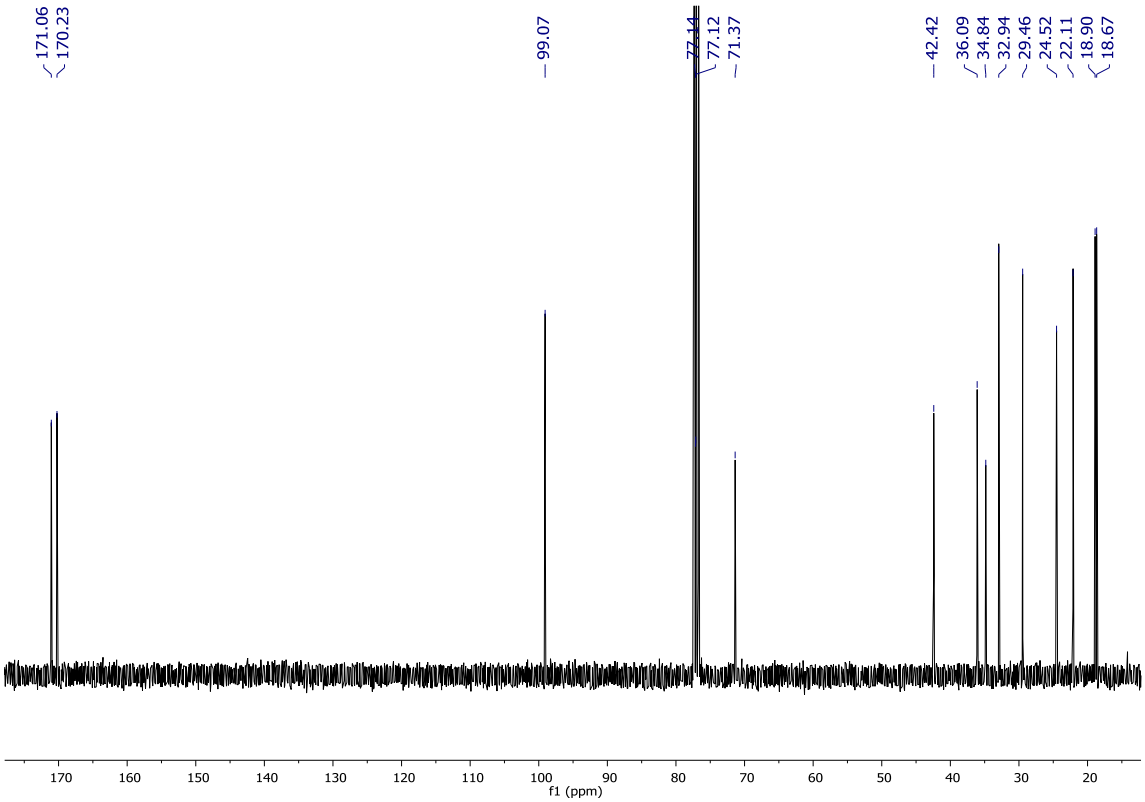


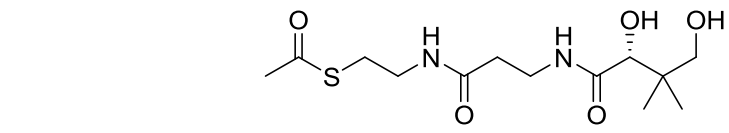
58





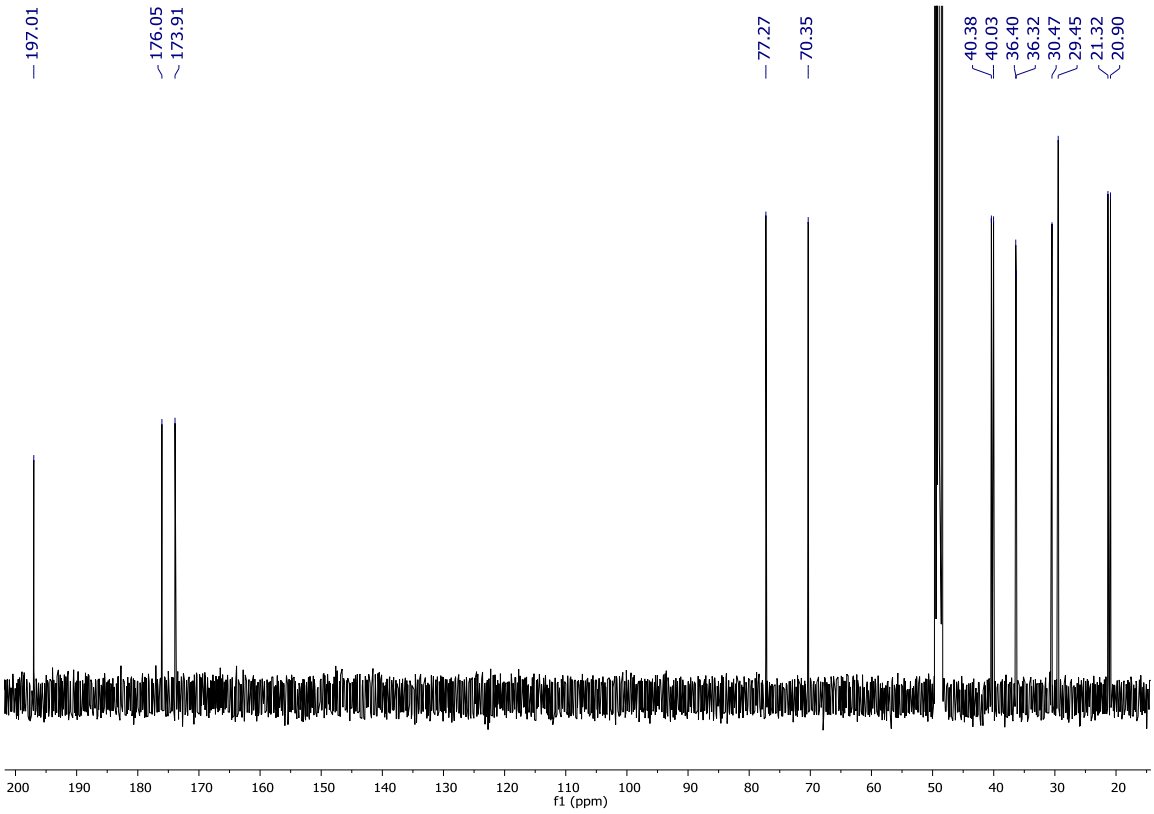

HS
H
N
O
H
N
O
O
O
59
0.92
0.94
1.10
1.03
2.12
2.06
1.03
1.97
1.98
2.85
2.83
1.06
3.00
2.96
f1 (ppm)

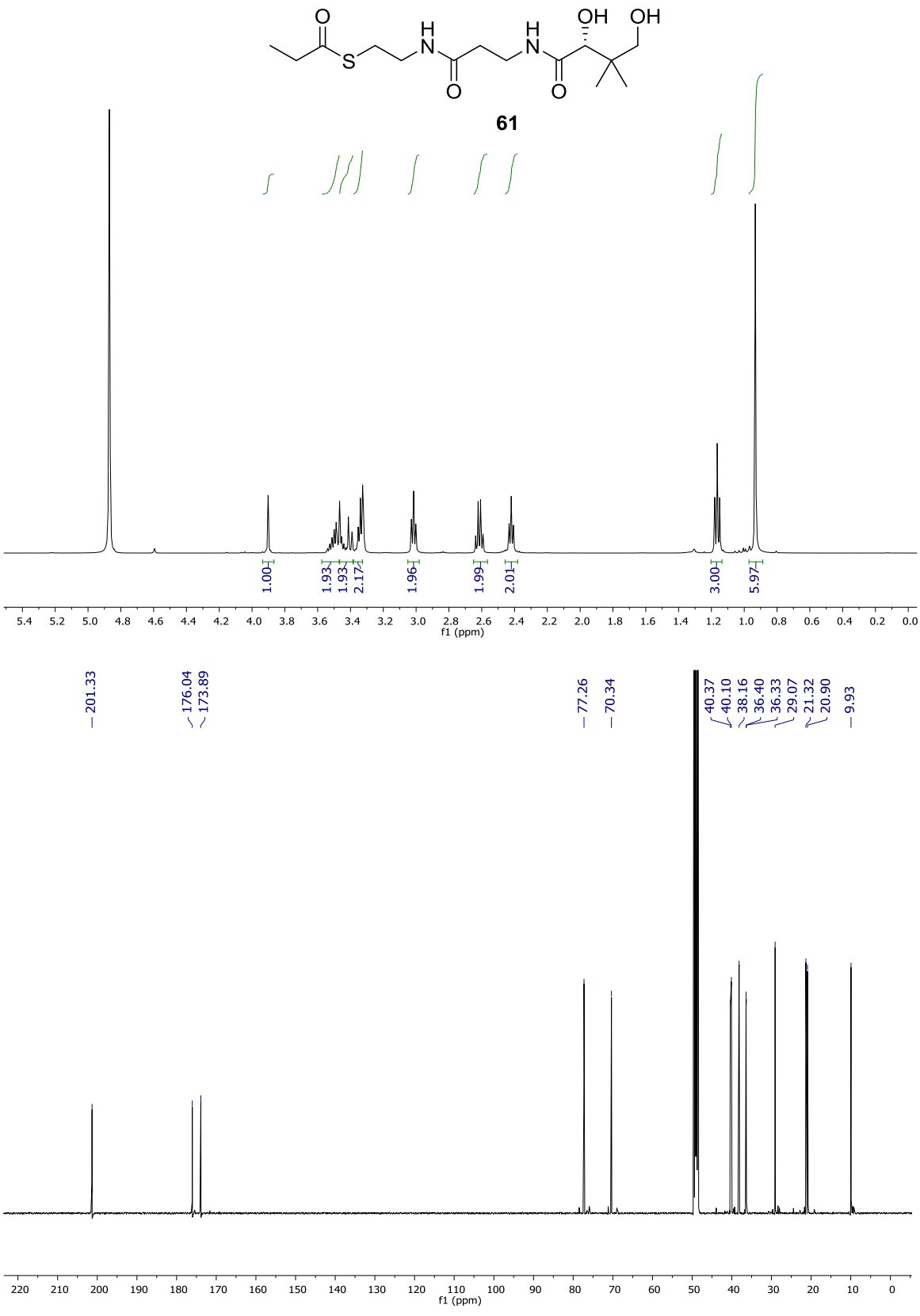

171.06
170.23
99.07
77.14
77.12
71.37
42.42
36.09
34.84
32.94
29.46
24.52
22.11
18.90
18.67
f1 (ppm)

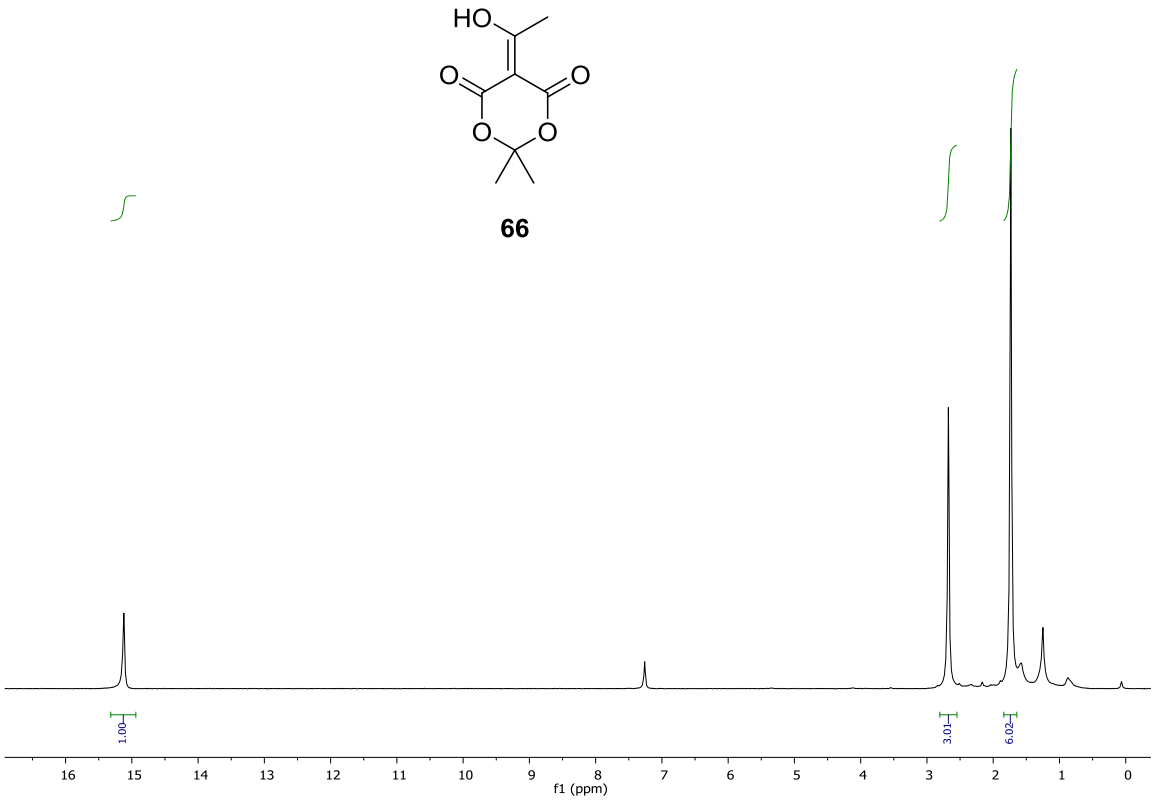


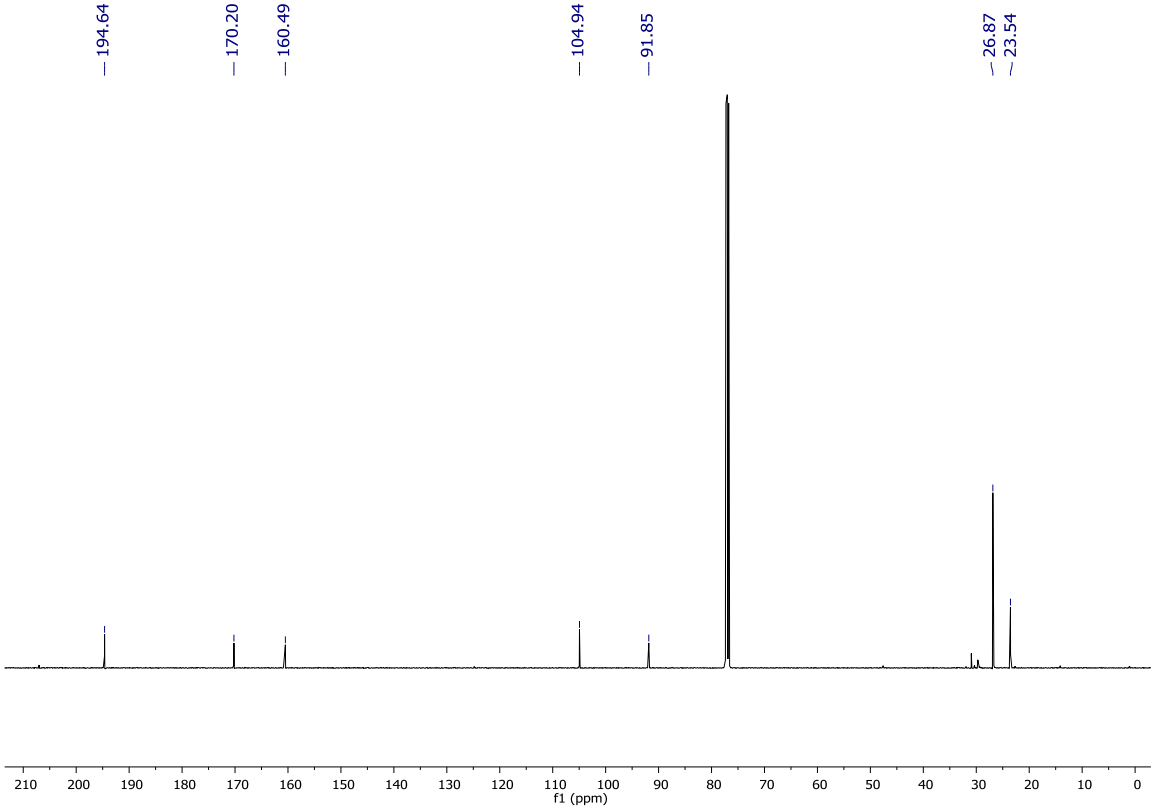
60

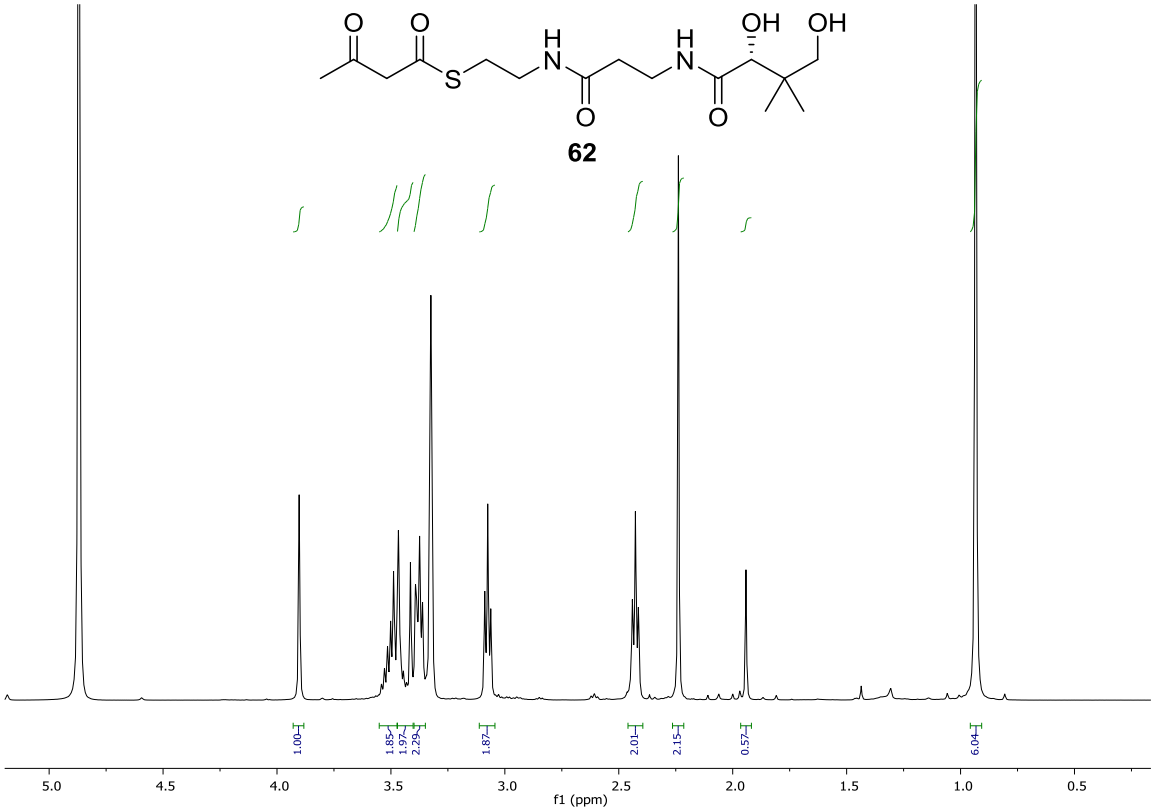
1.07
2.30
2.11
2.23
2.13
2.11
3.00
6.19
f1 (ppm)

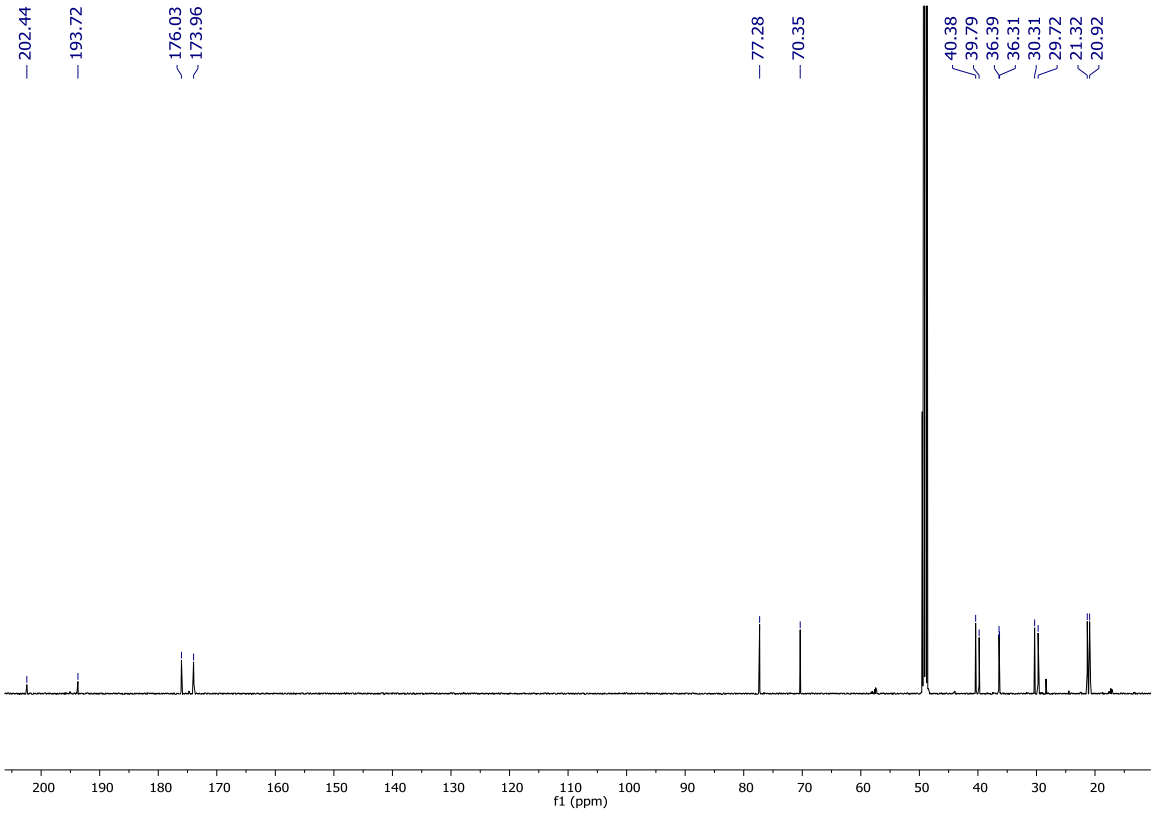

197.01
176.05
173.91
77.27
70.35
40.38
40.03
36.40
36.32
30.47
29.45
21.32
20.90
f1 (ppm)

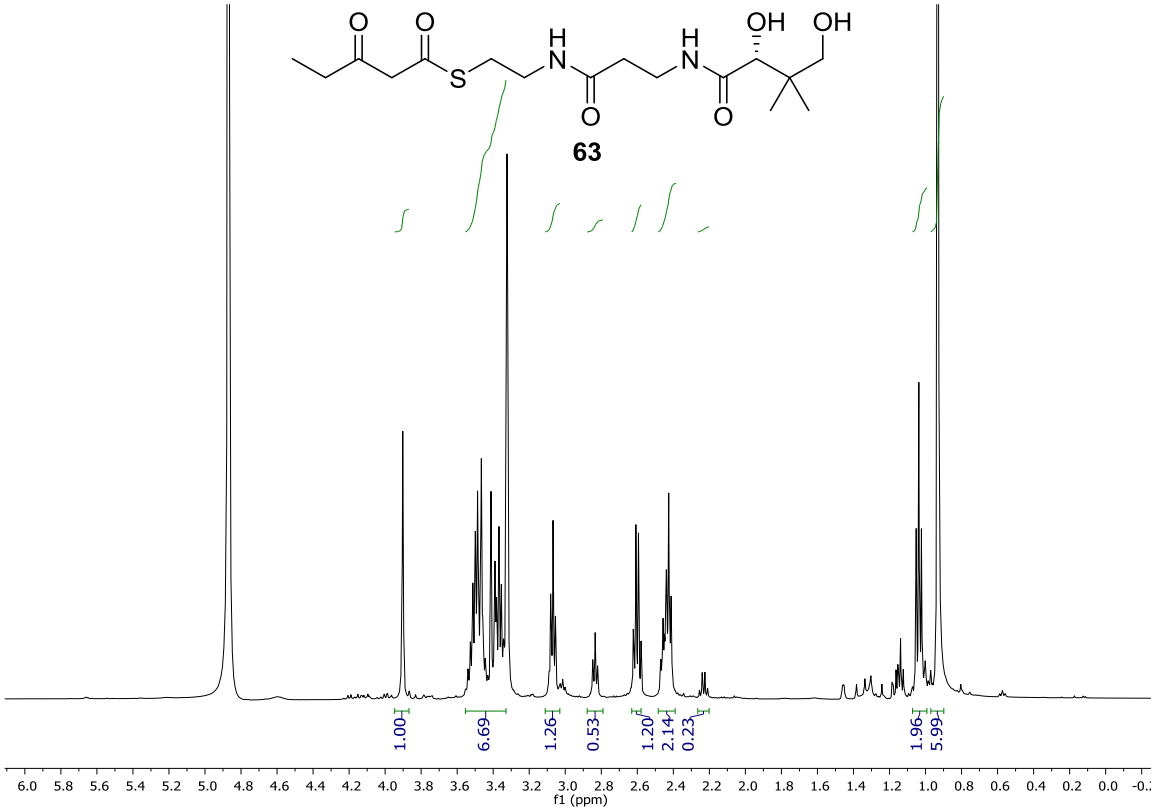

61
1.00
1.93
1.93
2.17
1.96
1.99
2.01
3.00
5.97
f1 (ppm)
201.33
176.04
173.89
77.26
70.34
40.37
40.10
38.16
36.40
36.33
29.07
21.32
20.90
9.93
f1 (ppm)

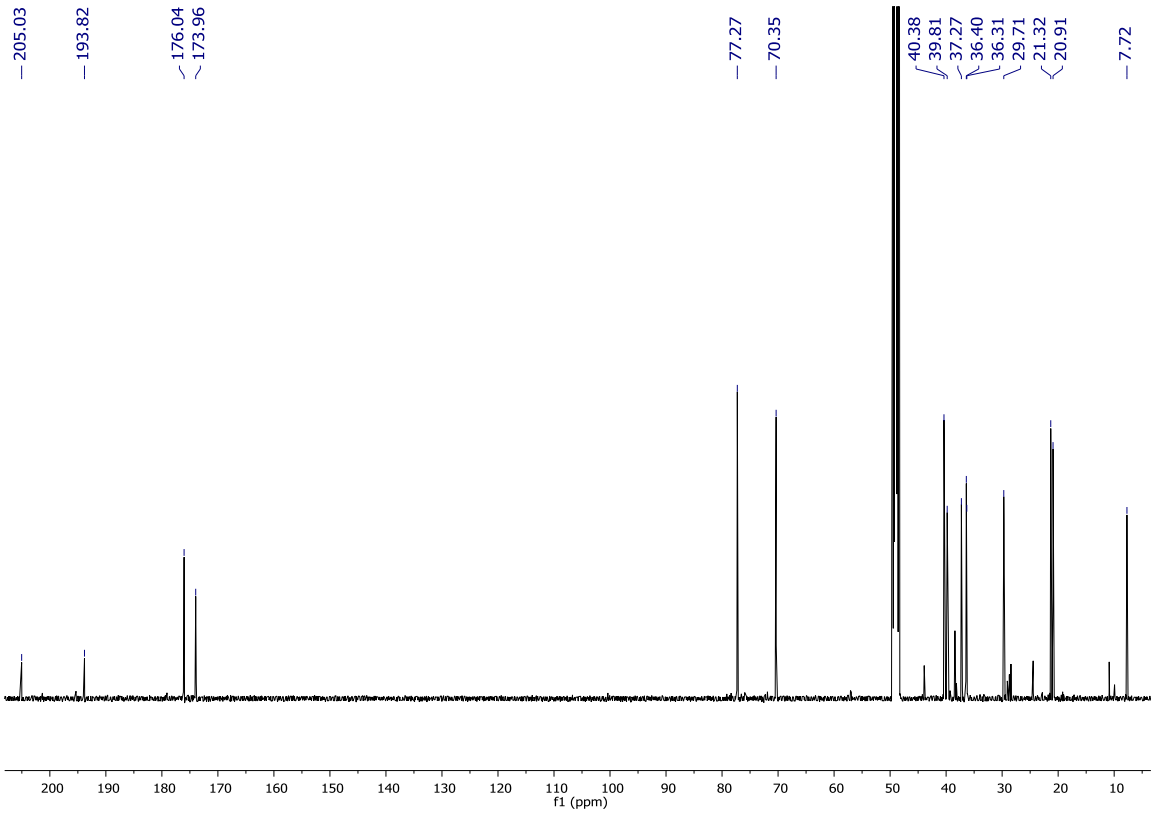

HO
O
O
O
O
66
1.00
3.01
6.02
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
f1 (ppm)

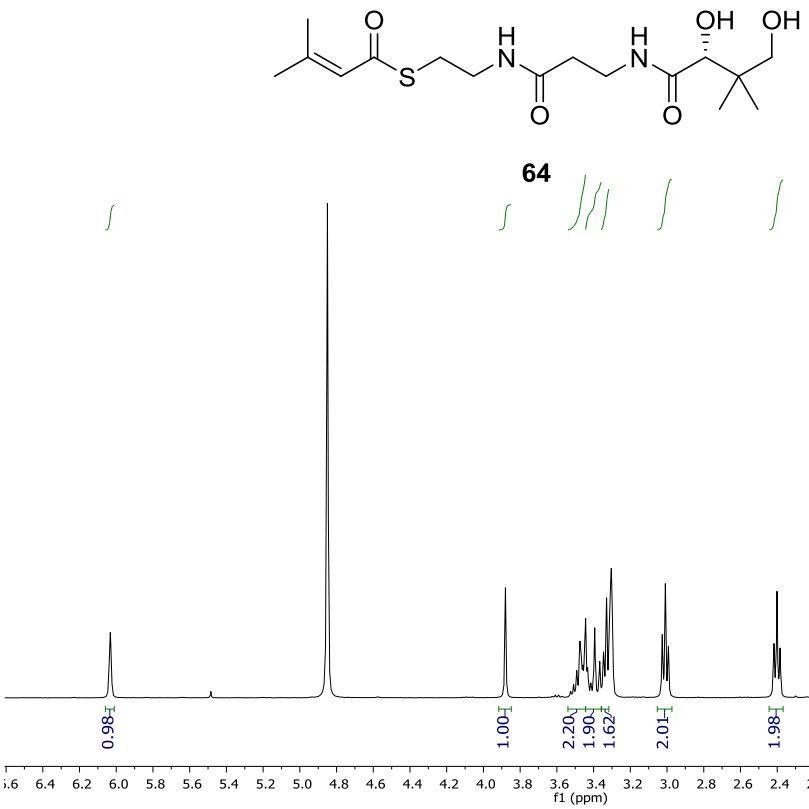

194.64
170.20
160.49
104.94
91.85
26.87
23.54
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

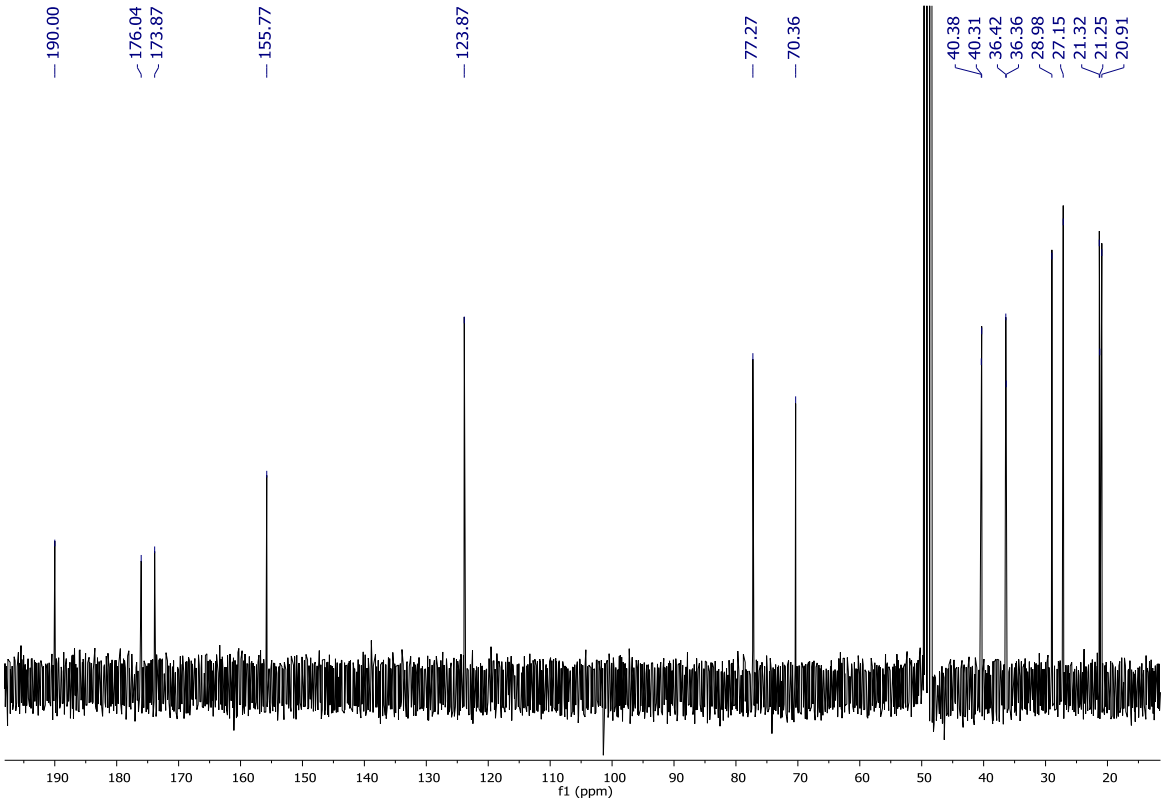

O
O
S
H
N
O
H
N
OH
OH
O
62
1.00
1.85
1.97
2.29
1.87
2.01
2.15
0.57
6.04
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

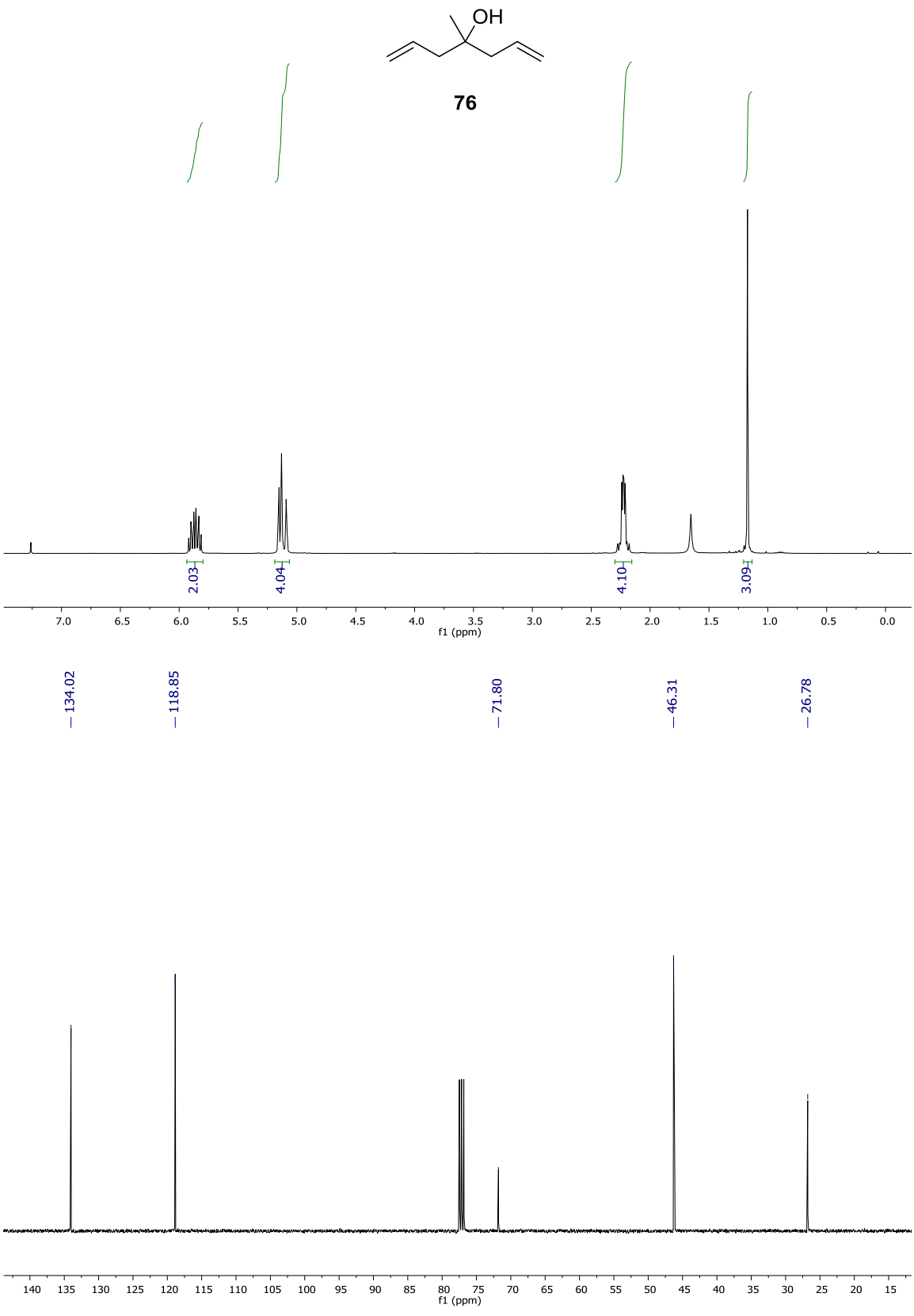

202.44
193.72
176.03
173.96
77.28
70.35
40.38
39.79
36.39
36.31
30.31
29.72
21.32
20.92
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
f1 (ppm)

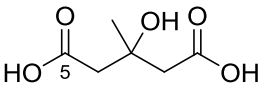

63
1.00
6.69
1.26
0.53
1.20
2.14
0.23
1.96
5.99
f1 (ppm)


205.03
193.82
176.04
173.96
77.27
70.35
40.38
39.81
37.27
36.40
36.31
29.71
21.32
20.91
7.72
f1 (ppm)

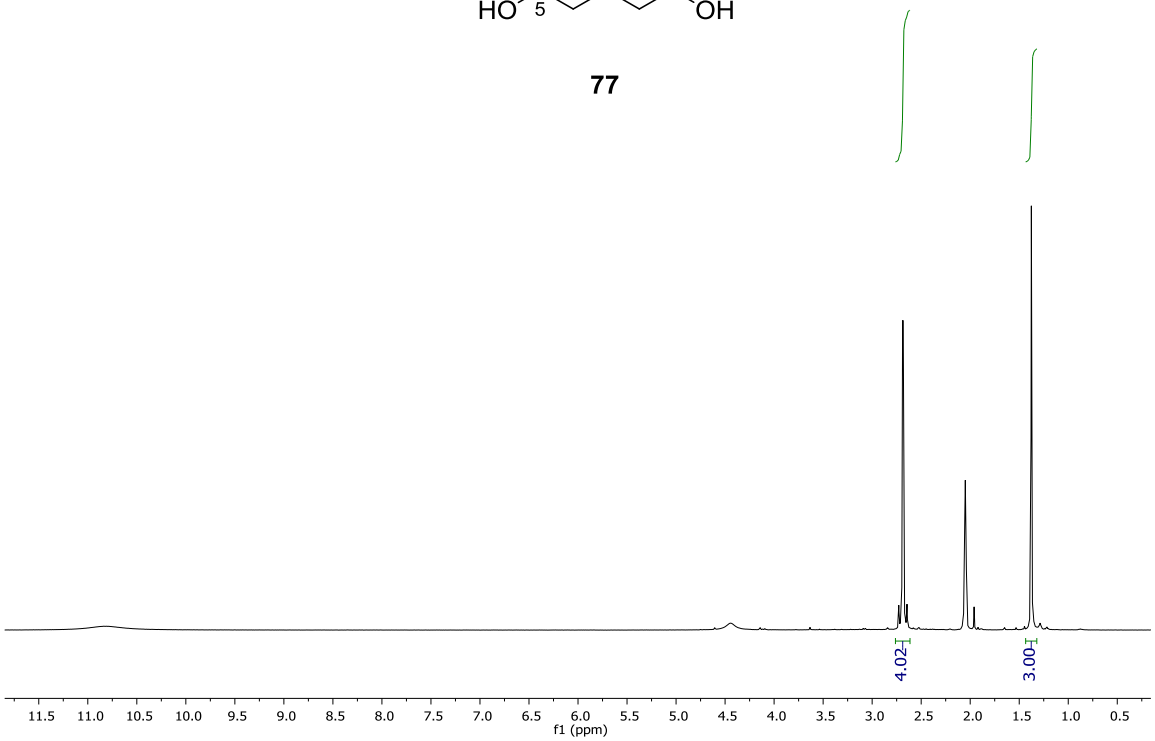

O
S
H
N
O
H
N
O
OH
OH
64
0.98
1.00
2.20
1.90
1.62
2.01
1.98
3.00
2.99
6.02
f1 (ppm)

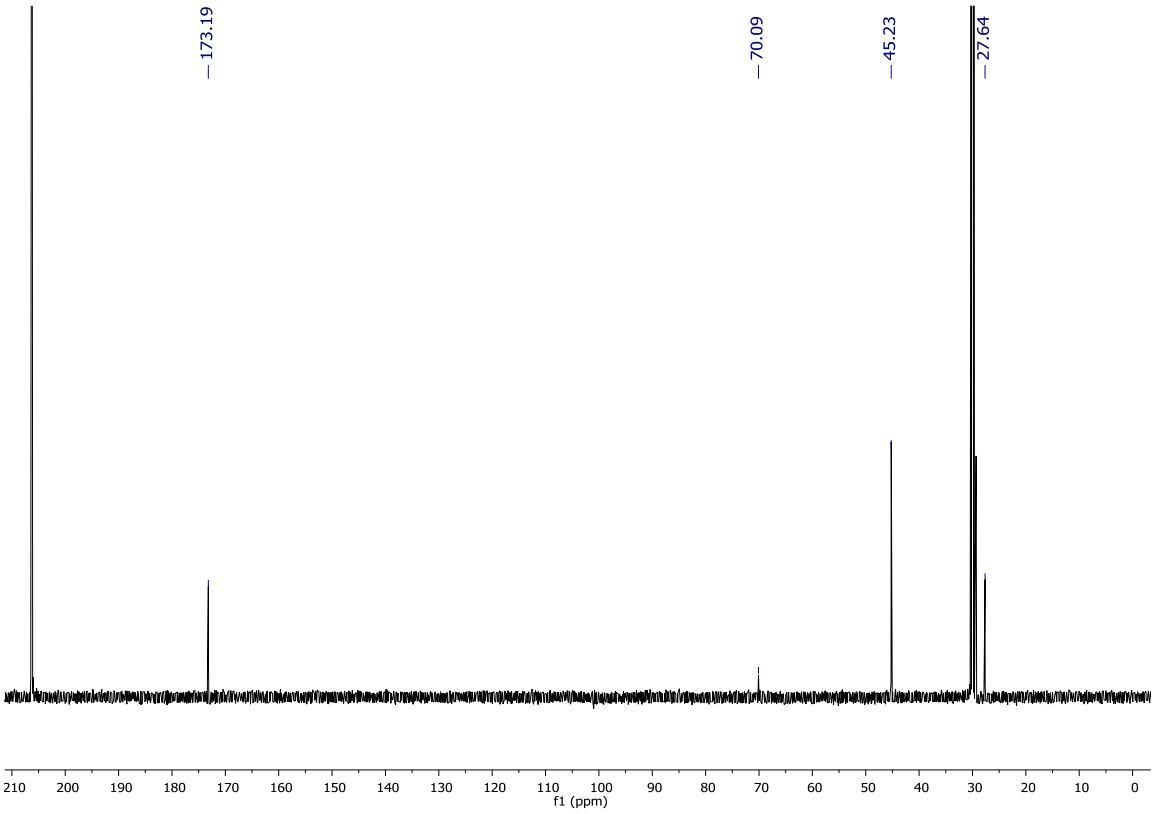

190.00
176.04
173.87
155.77
123.87
77.27
70.36
40.38
40.31
36.42
36.36
28.98
27.15
21.32
21.25
20.91
f1 (ppm)

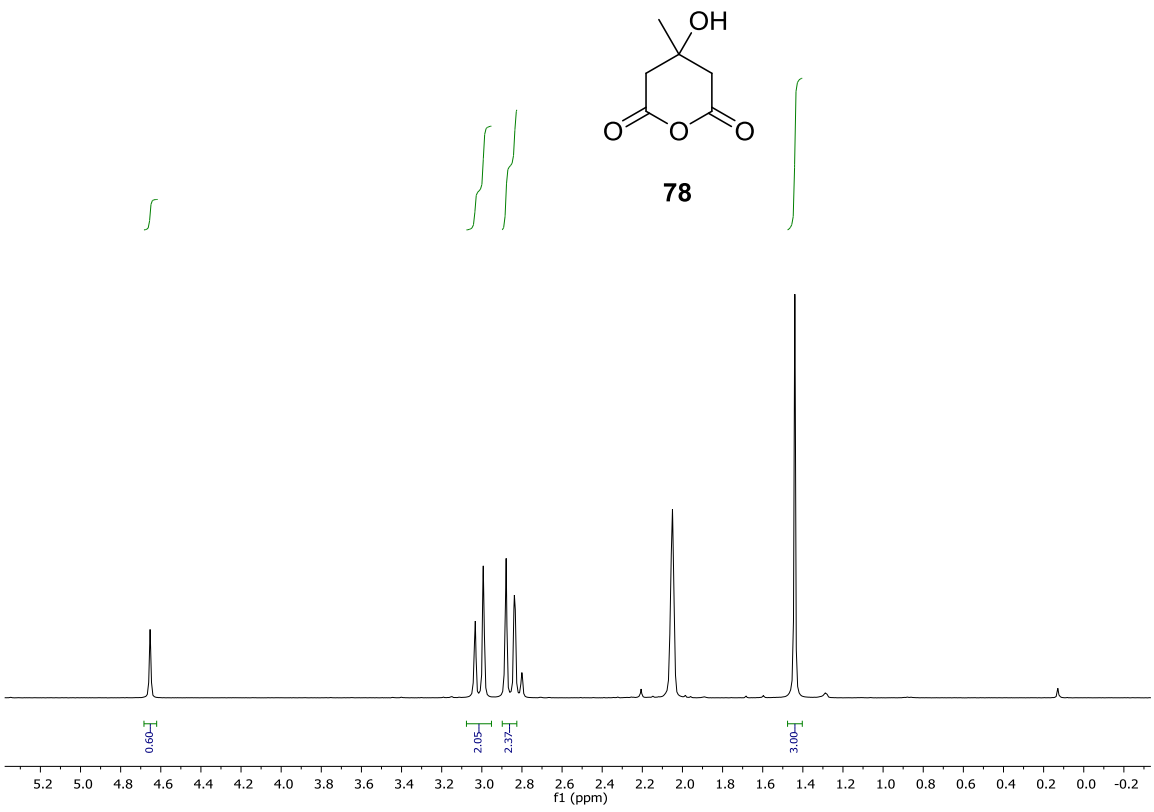

OH
76
2.03
4.04
4.10
3.09
f1 (ppm)
134.02
118.85
71.80
46.31
26.78
f1 (ppm)

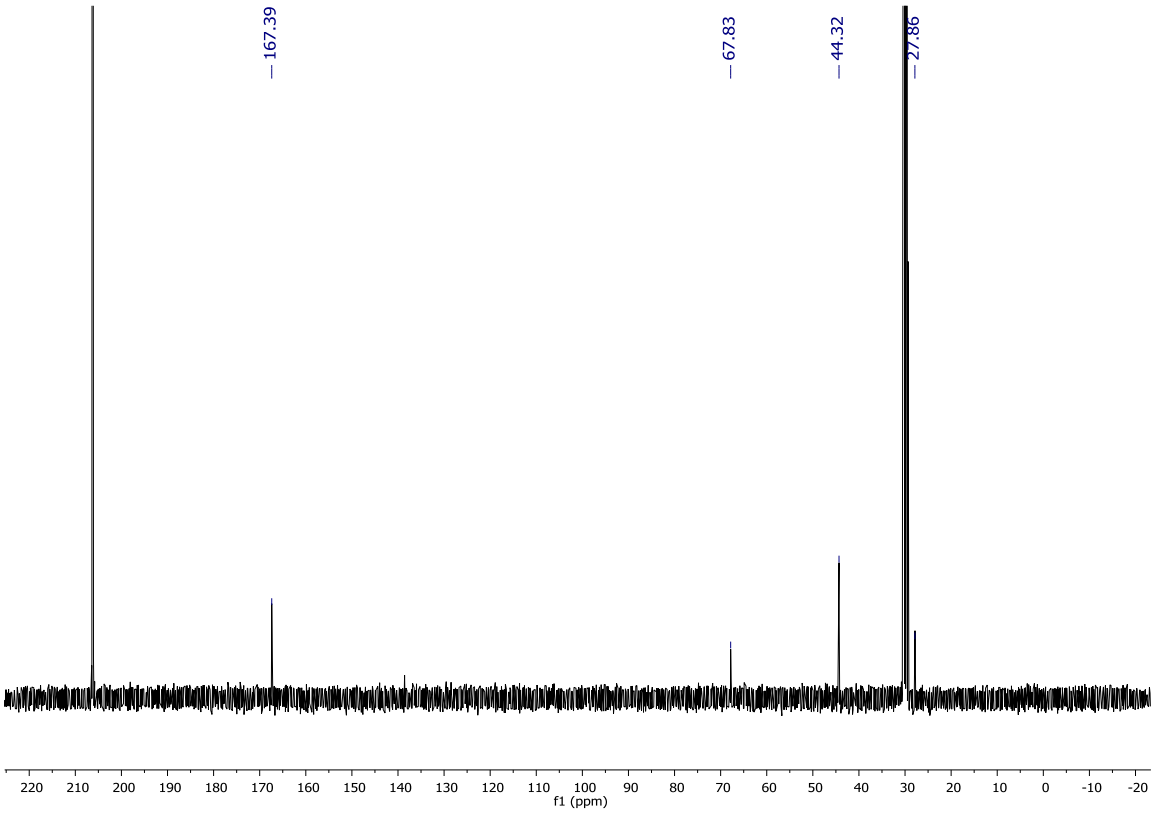

O
OH
O
HO
5
OH

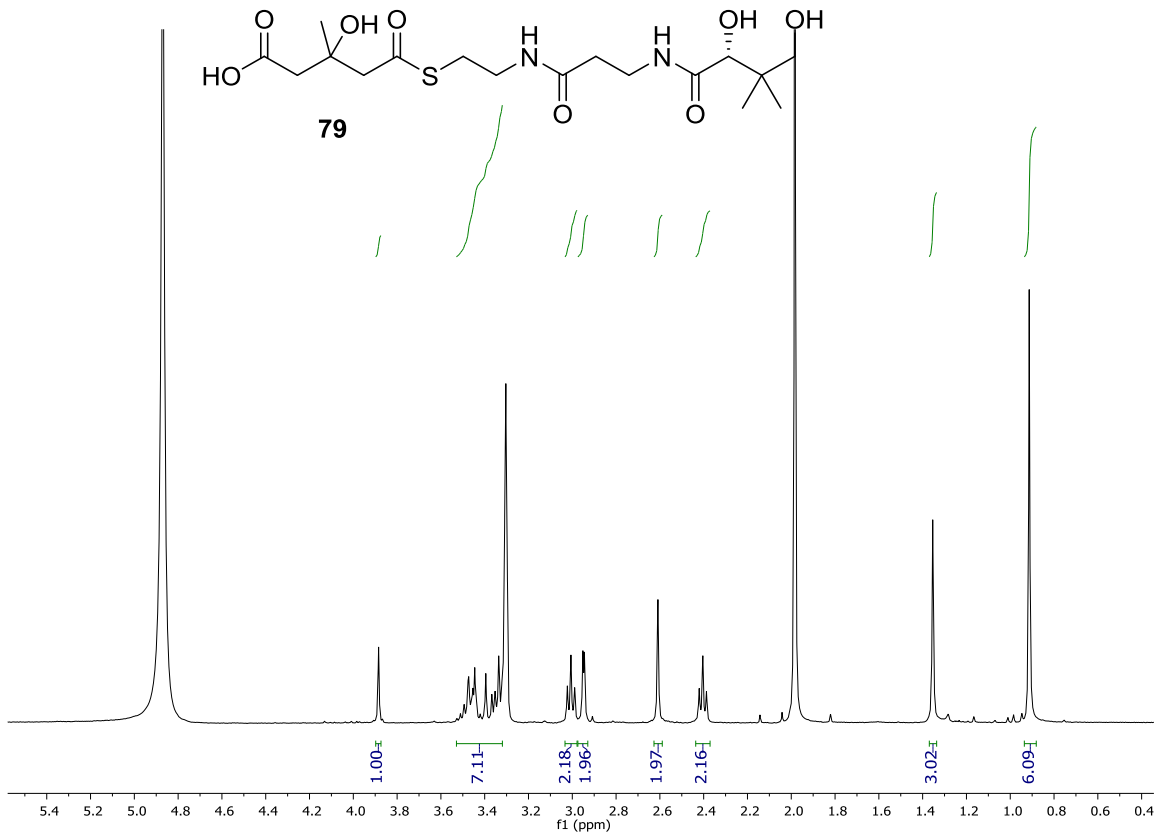
77

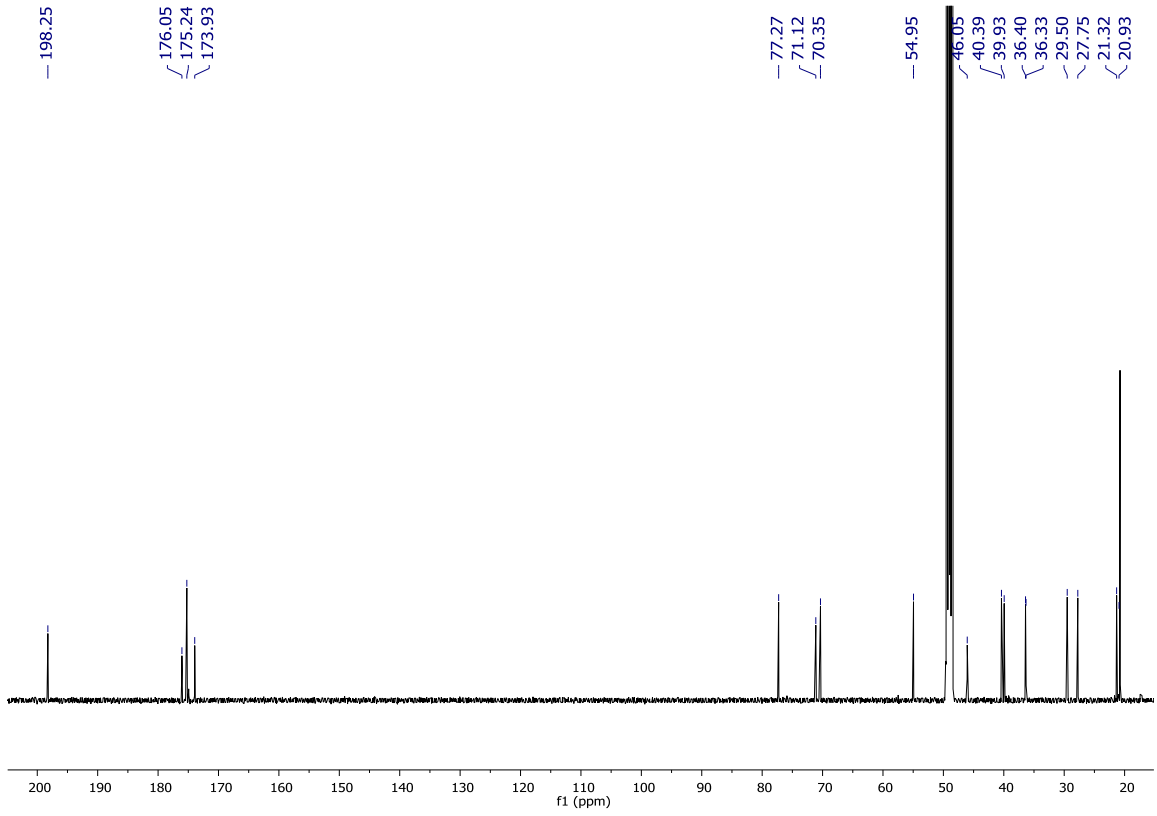

4.02
3.00
11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
f1 (ppm)

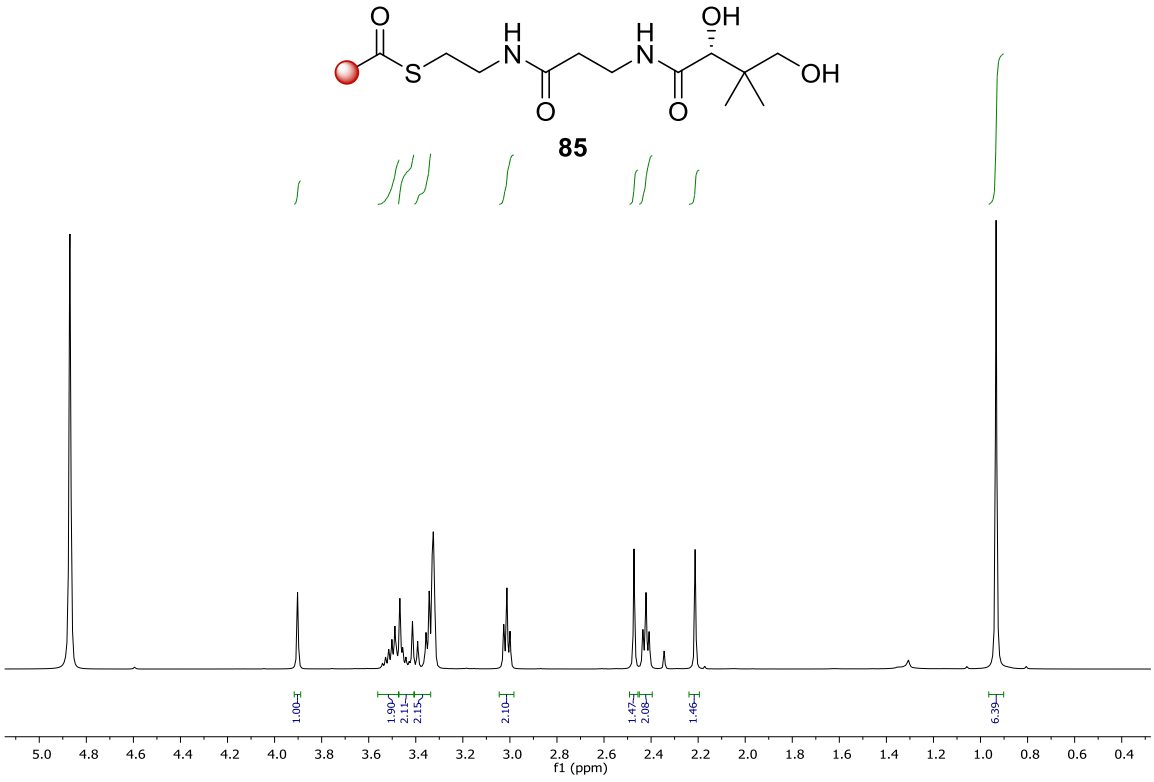

173.19
70.09
45.23
27.64
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

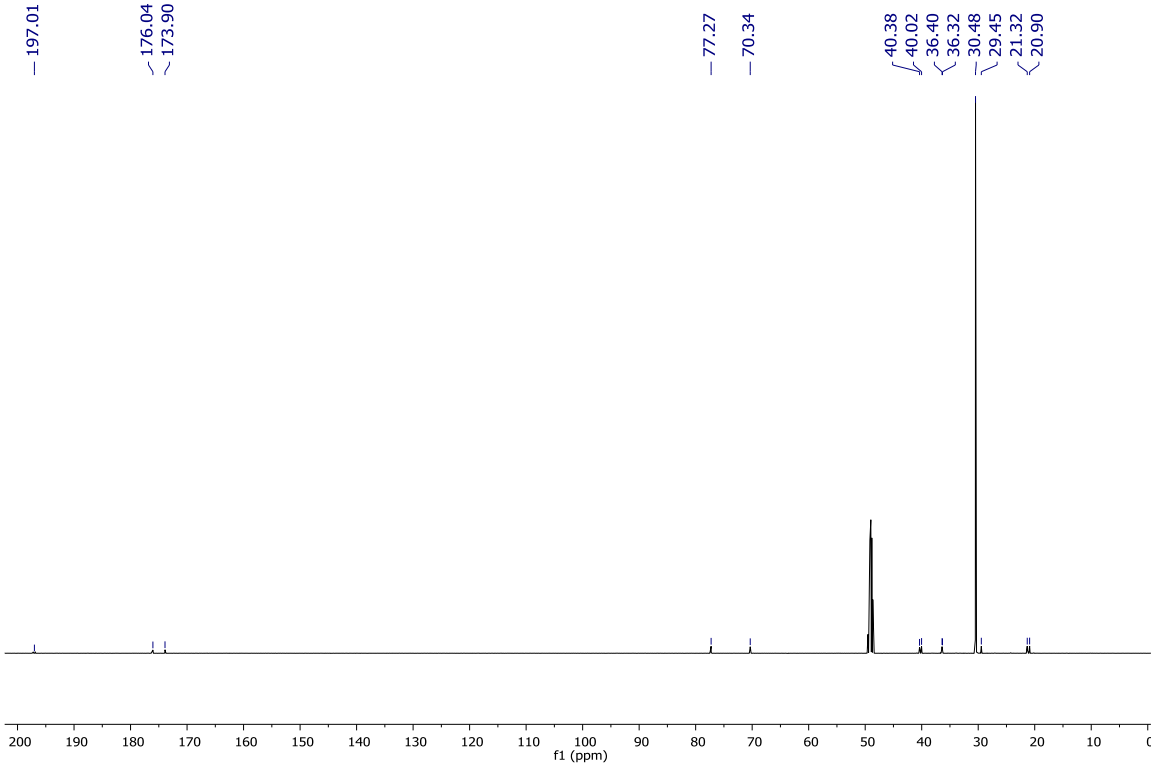

OH
O
O
O
78
0.60
2.05
2.37
3.00
f1 (ppm)

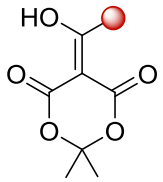

167.39
67.83
44.32
27.86
f1 (ppm)


79
1.00
7.11
2.18
1.96
1.97
2.16
3.02
6.09
f1 (ppm)


198.25
176.05
175.24
173.93
77.27
71.12
70.35
54.95
46.05
40.39
39.93
36.40
36.33
29.50
27.75
21.32
20.93
f1 (ppm)

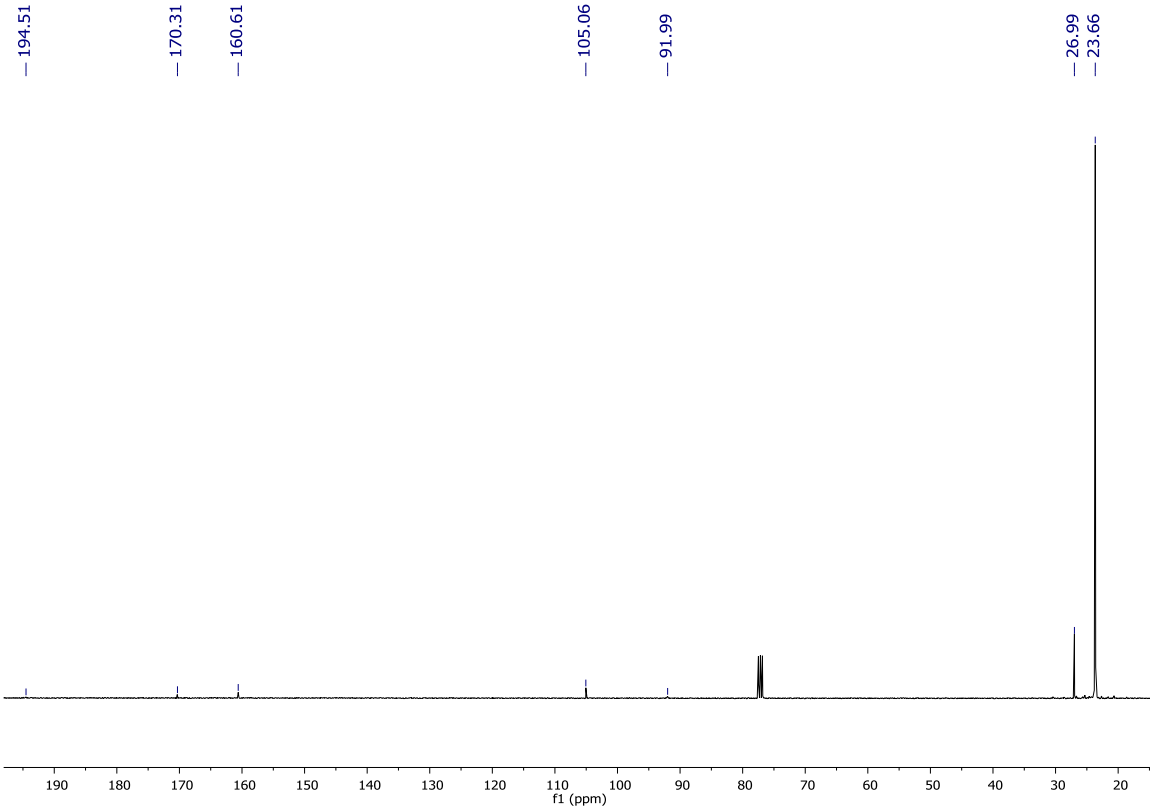

85
1.00
1.90
2.11
2.15
2.10
1.47
2.08
1.46
6.39
f1 (ppm)

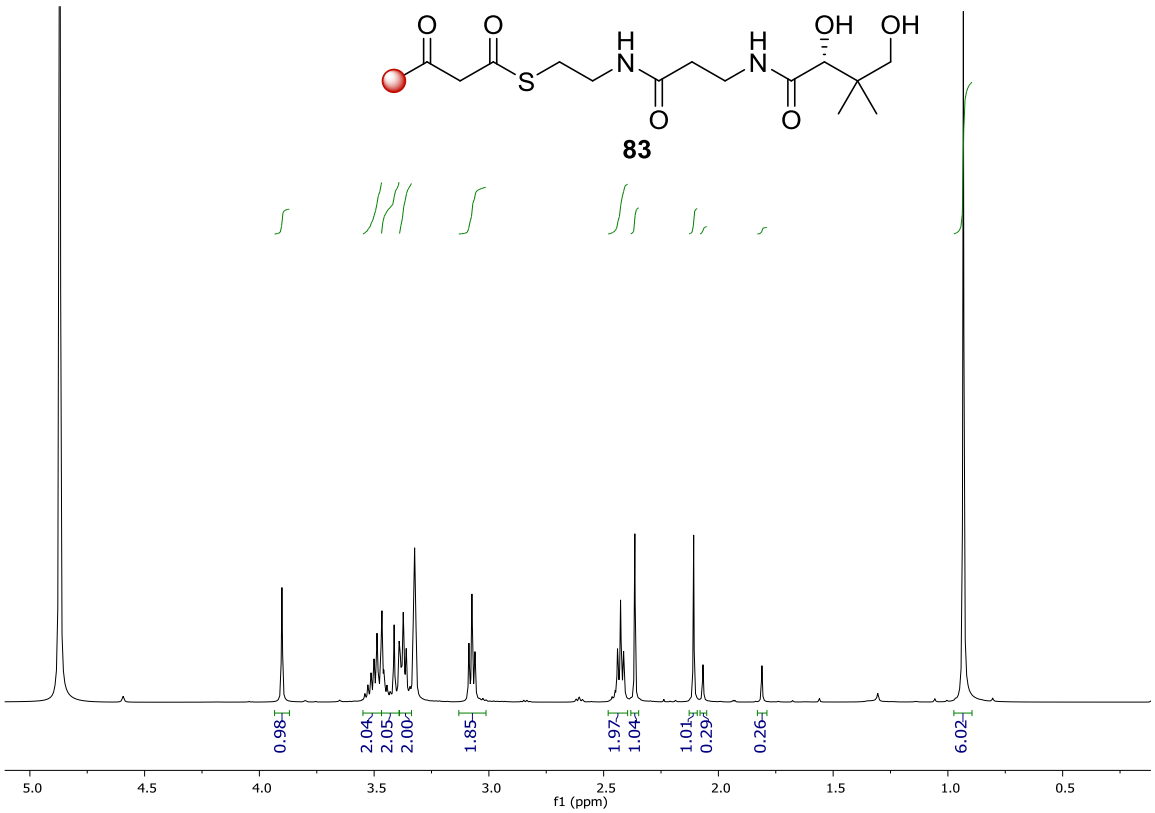

197.01
176.04
173.90
77.27
70.34
40.38
40.02
36.40
36.32
30.48
29.45
21.32
20.90
f1 (ppm)

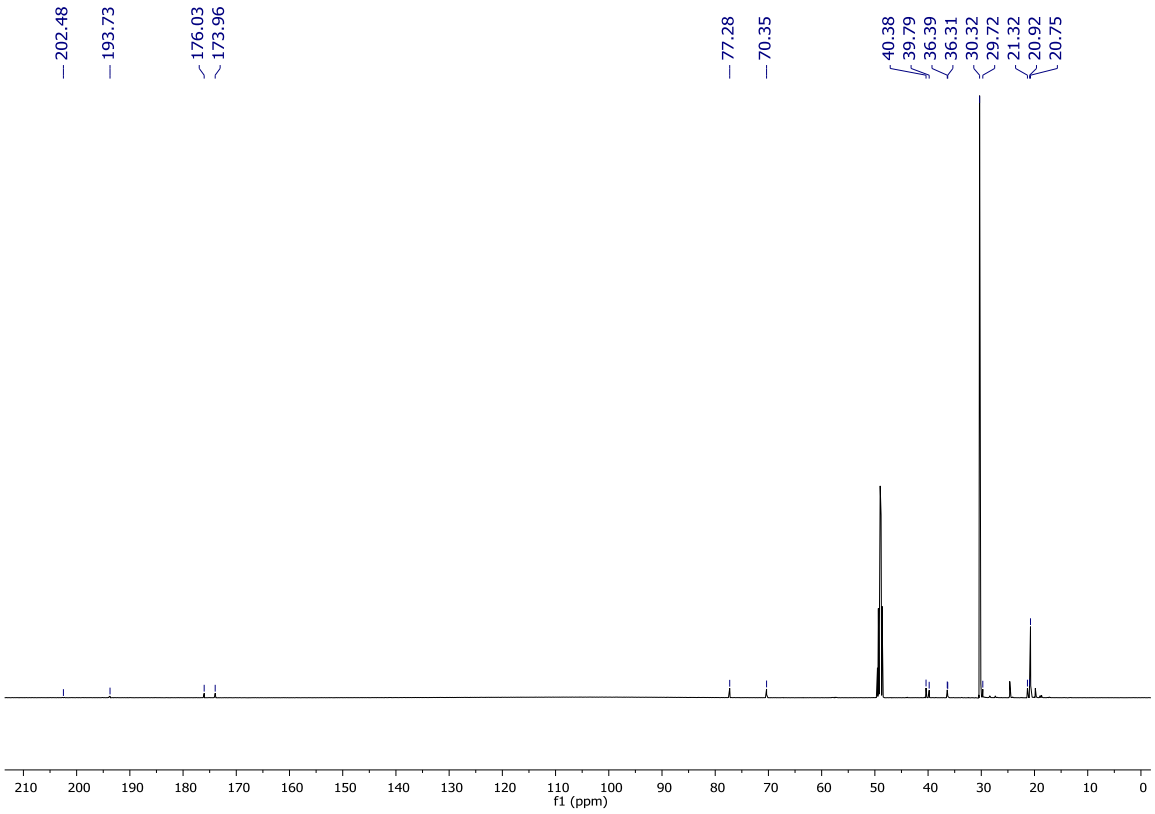

HO
O
O
O
O

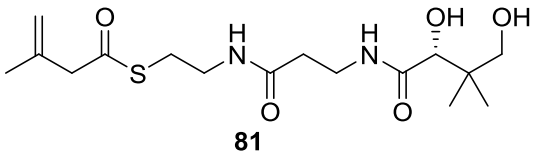
87

15.5 15.0 14.5 14.0 13.5 13.0 12.5 12.0 11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5
f1 (ppm)


194.51
170.31
160.61
105.06
91.99
26.99
23.66
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20
f1 (ppm)

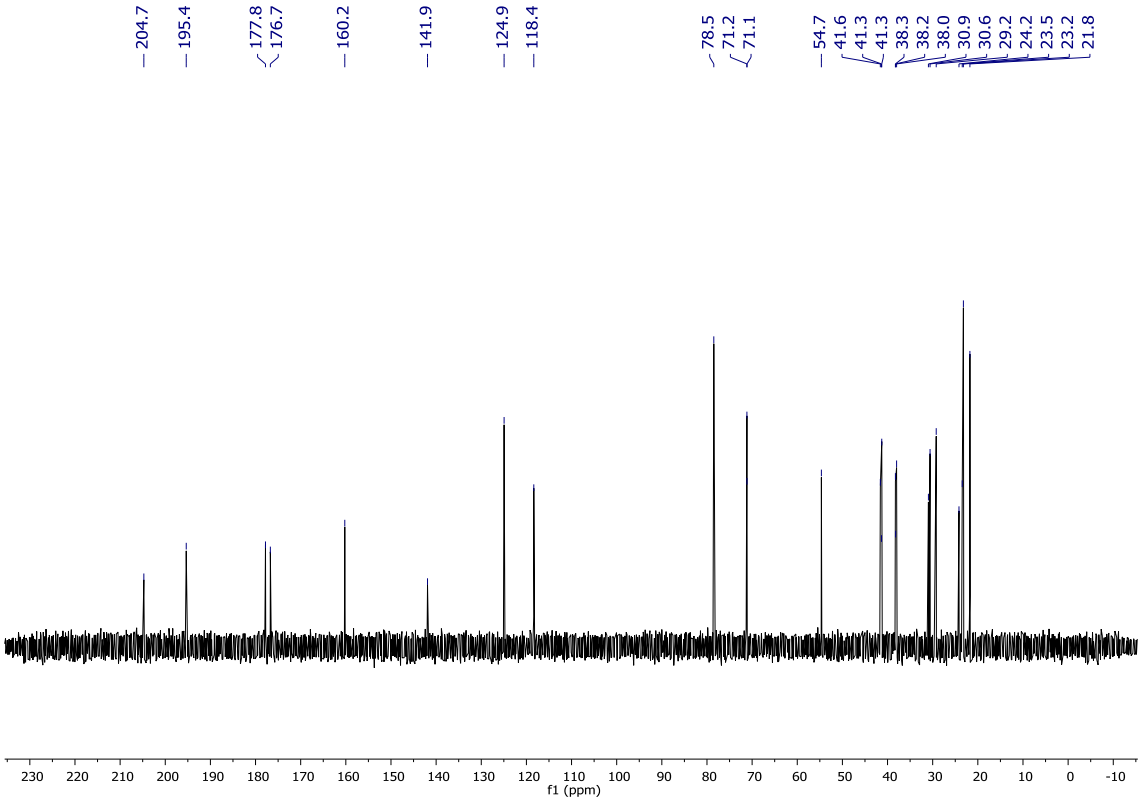

O
O
S
H
N
O
H
N
OH
OH
O
83
0.98
2.04
2.05
2.00
1.85
1.97
1.04
1.01
0.29
0.26
6.02
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

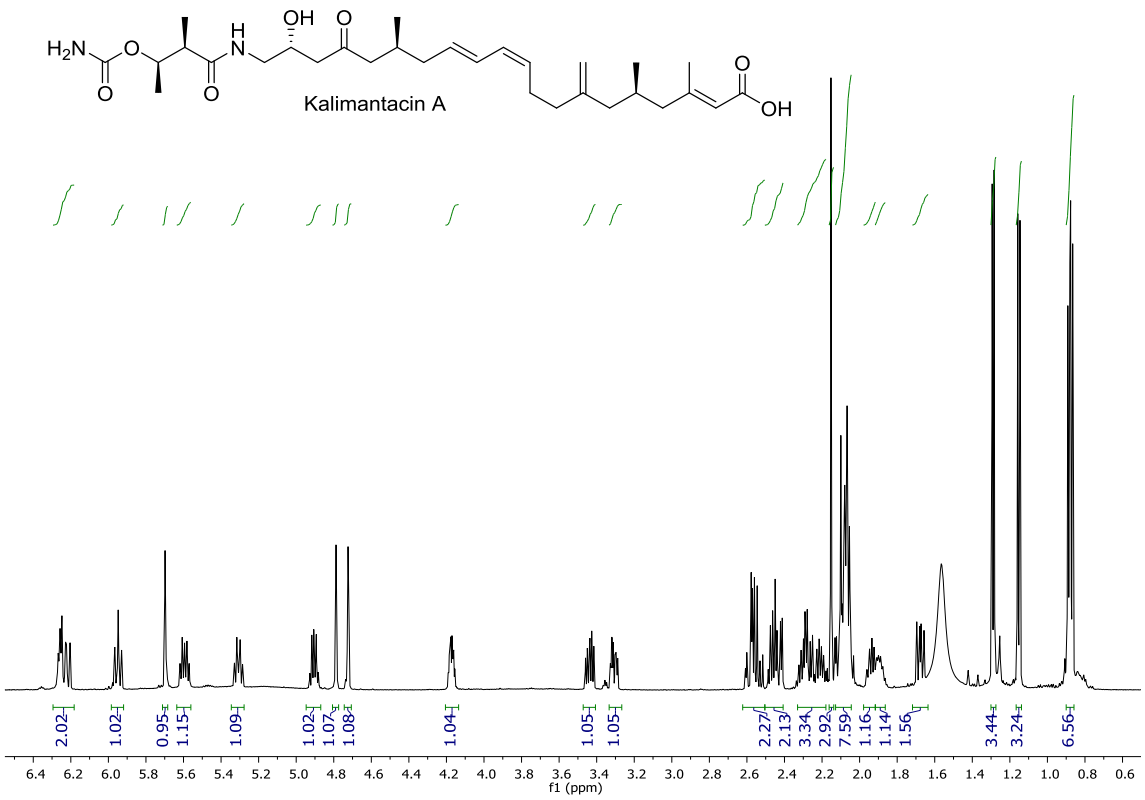

202.48
193.73
176.03
173.96
77.28
70.35
40.38
39.79
36.39
36.31
30.32
29.72
21.32
20.92
20.75
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

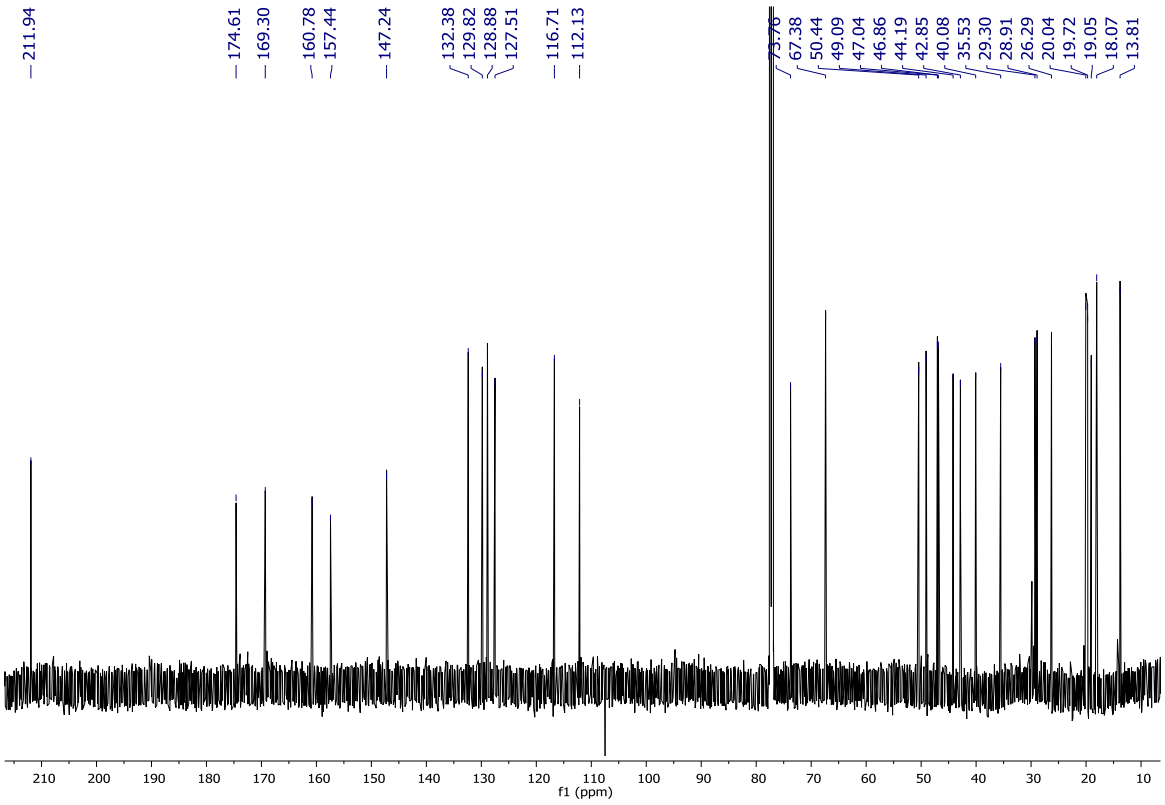

O
S
H
N
O
H
N
OH
OH
O

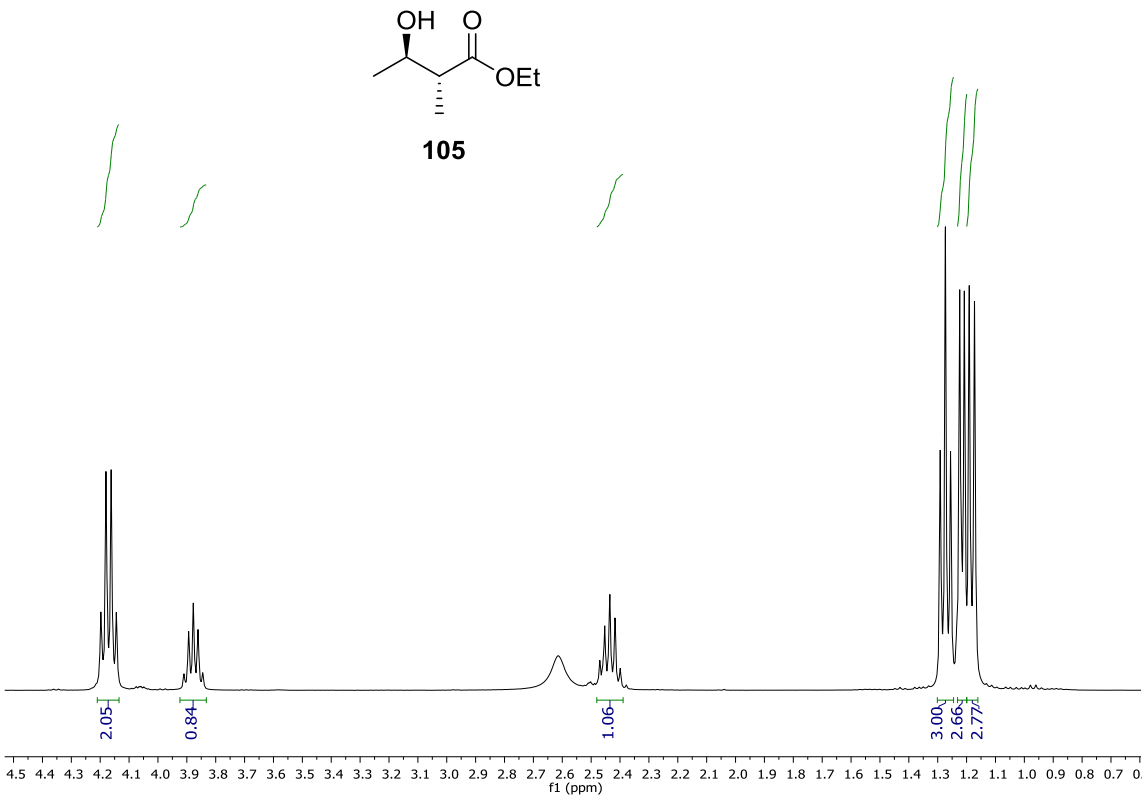
81

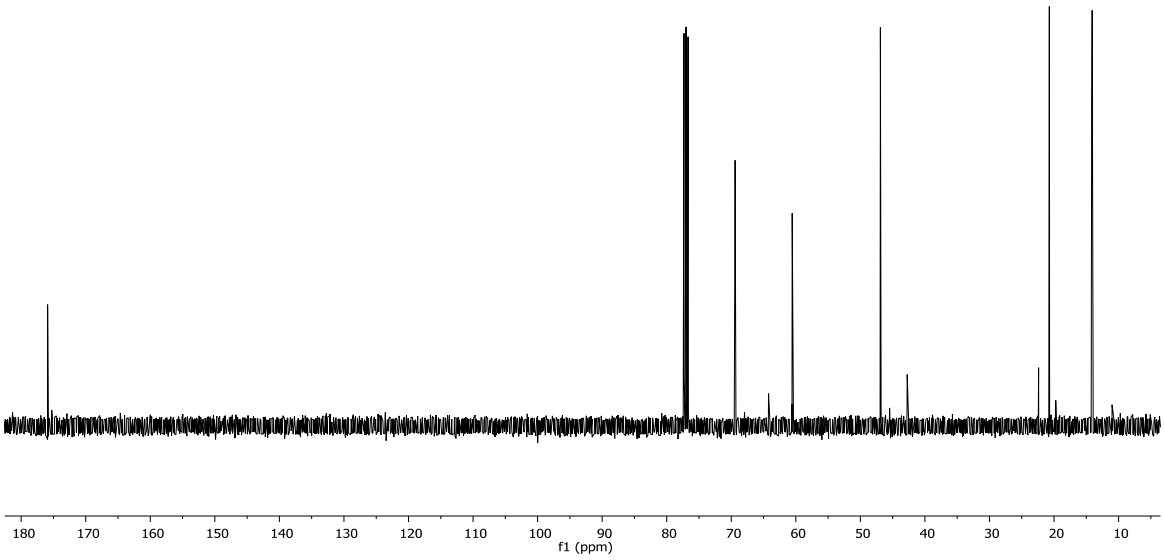
0.31
0.37
0.36
0.04
0.81
3.62
3.84
2.00
2.12
1.33
1.63
0.78
3.17
3.12
f1 (ppm)

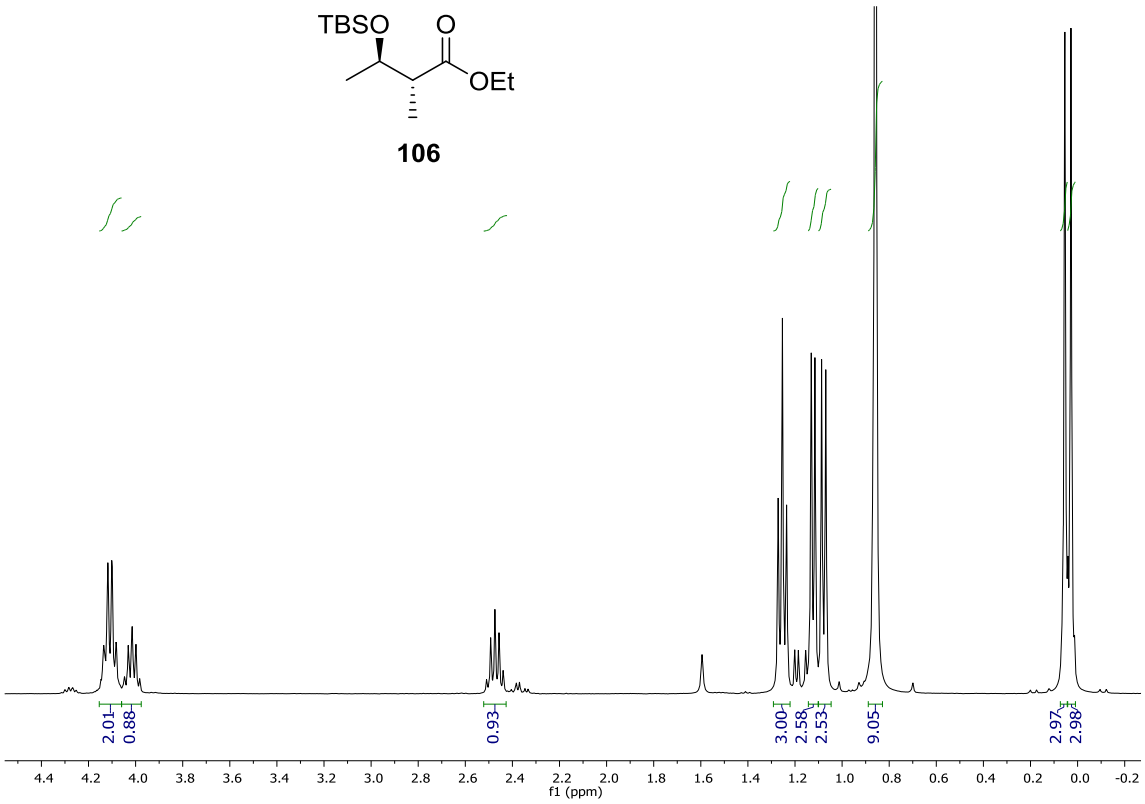

204.7
195.4
177.8
176.7
160.2
141.9
124.9
118.4
78.5
71.2
71.1
54.7
41.6
41.3
41.3
38.3
38.2
38.0
30.9
30.6
29.2
24.2
23.5
23.2
21.8
f1 (ppm)

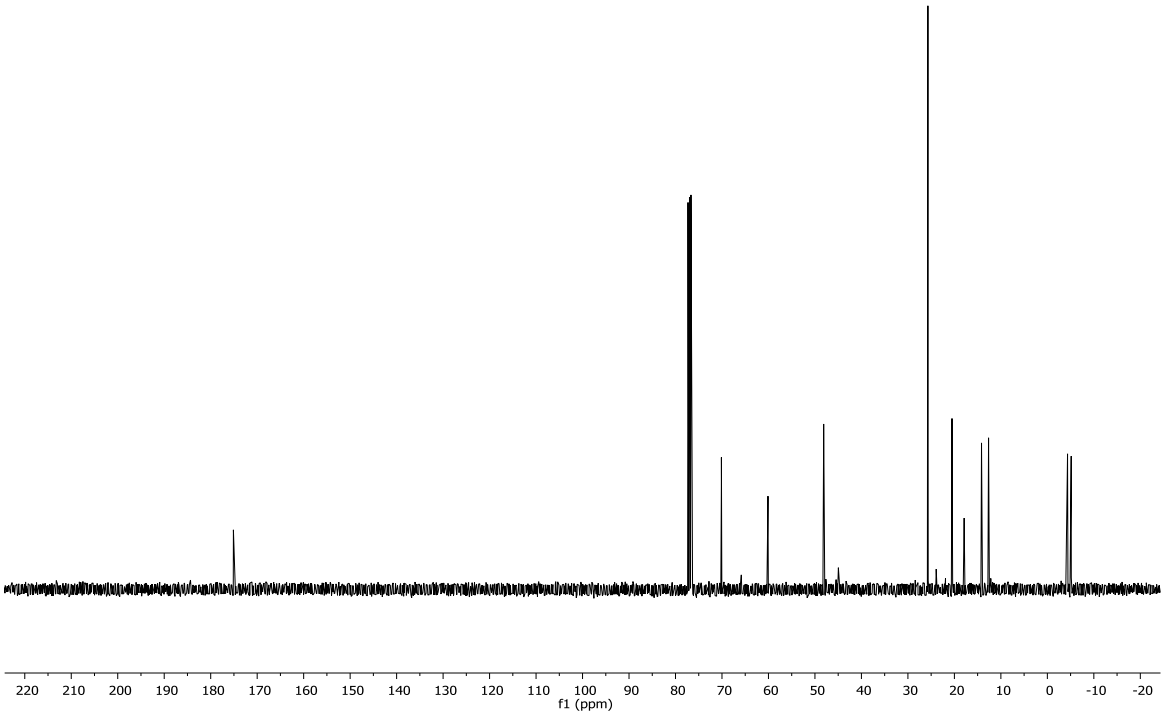

Kalimantacin A
2.02
1.02
0.95
1.15
1.09
1.02
1.07
1.08
1.04
1.05
1.05
2.27
2.13
3.34
2.92
7.59
1.16
1.14
1.56
3.44
3.24
6.56
f1 (ppm)

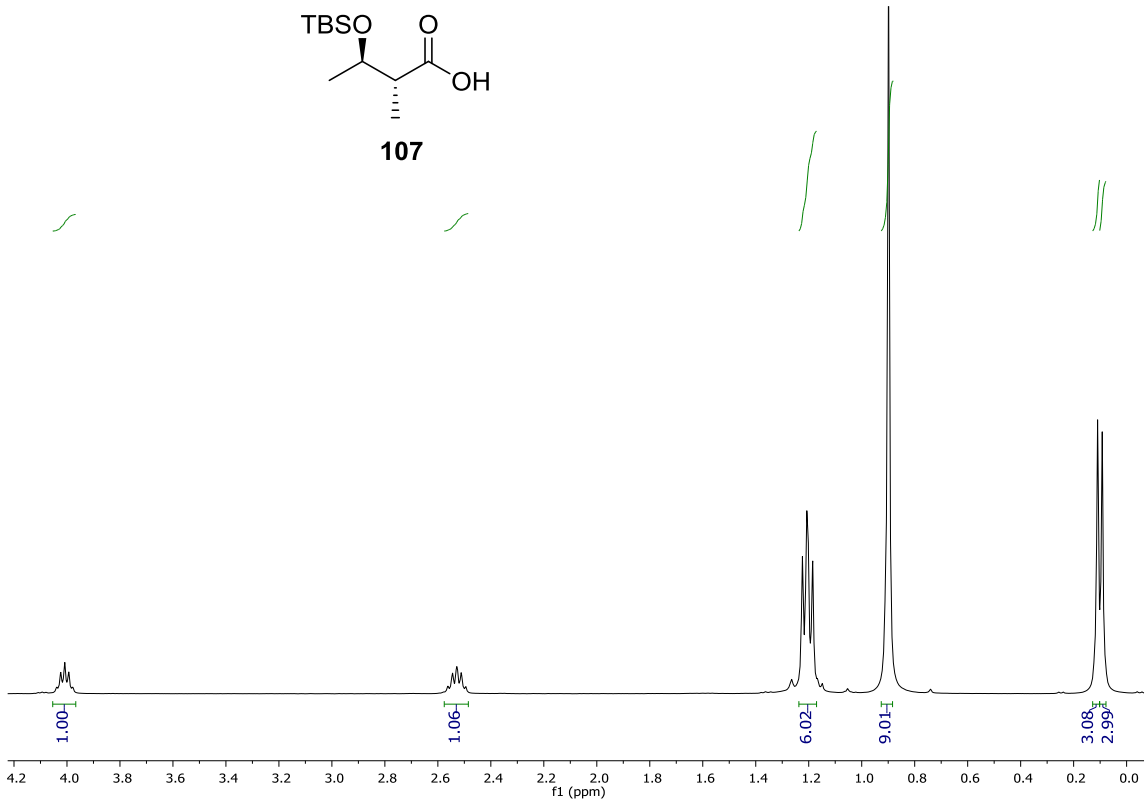

211.94
174.61
169.30
160.78
157.44
147.24
132.38
129.82
128.88
127.51
116.71
112.13
73.76
67.38
50.44
49.09
47.04
46.86
44.19
42.85
40.08
35.53
29.30
28.91
26.29
20.04
19.72
19.05
18.07
13.81
f1 (ppm)

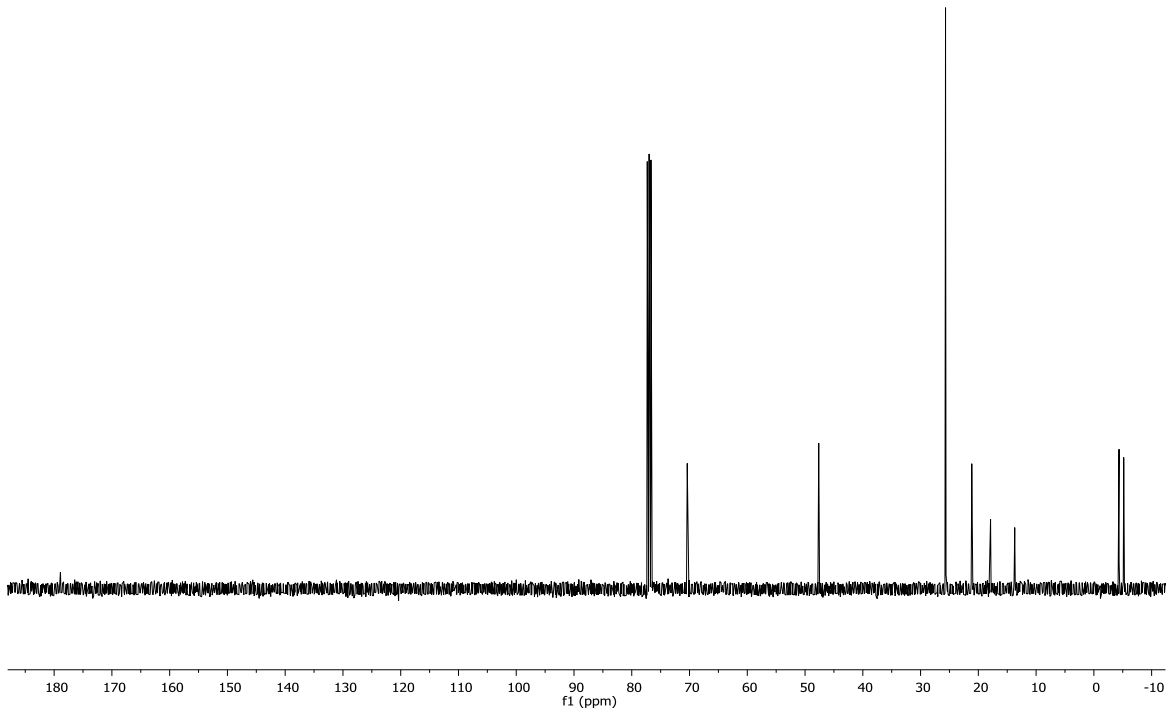

OH
O
OEt
105
2.05
0.84
1.06
3.00
2.66
2.77
f1 (ppm)

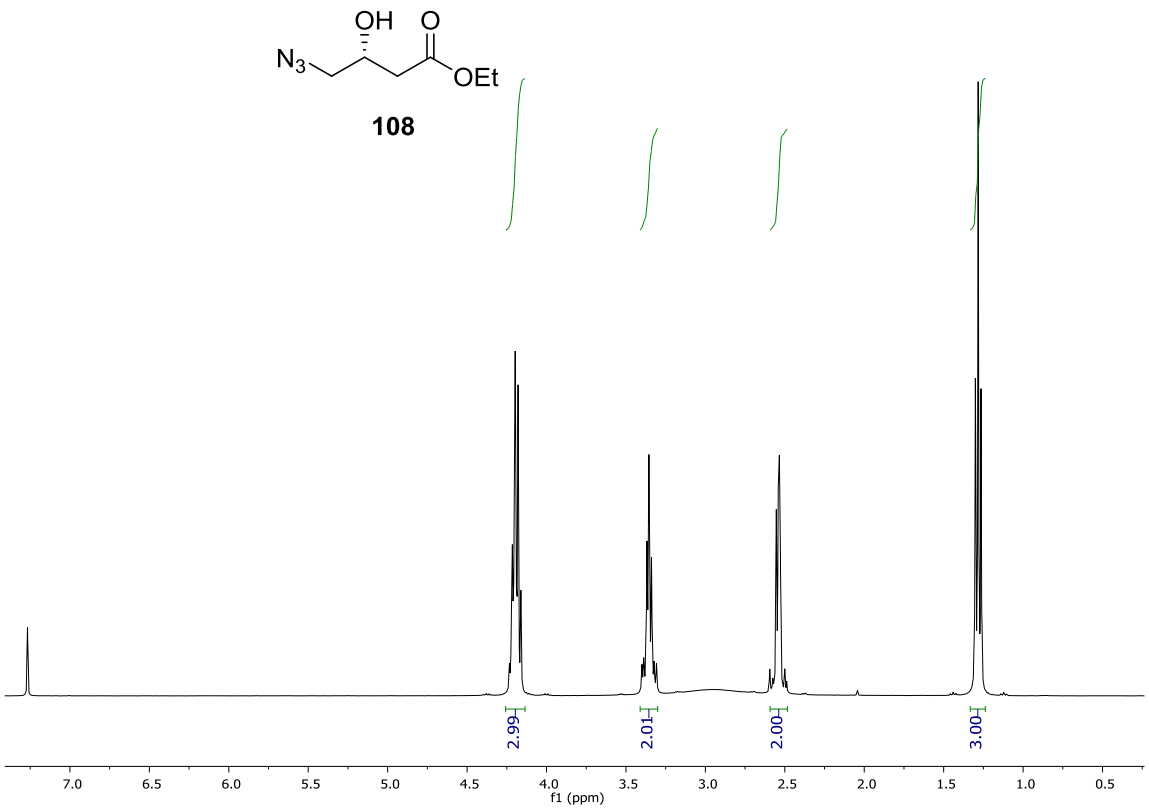

f1 (ppm)

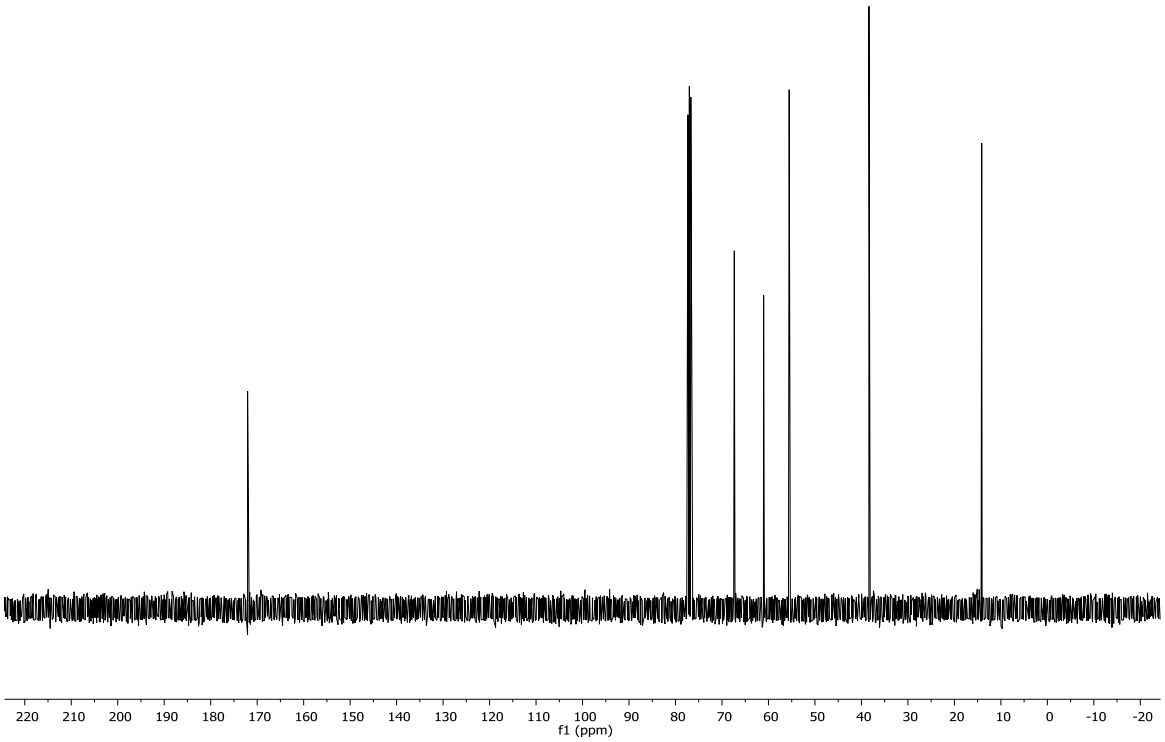

TBSO
O
OEt
106
2.01
0.88
0.93
3.00
2.58
2.53
9.05
2.97
2.98
f1 (ppm)

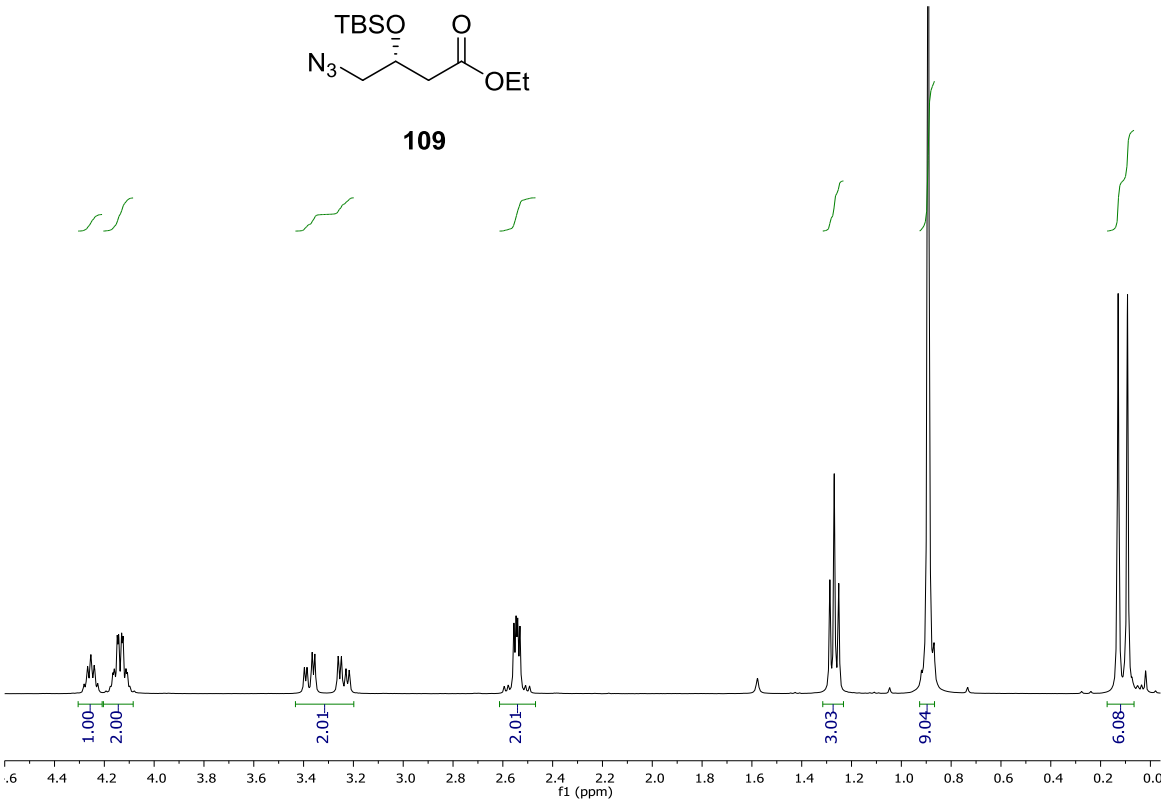

f1 (ppm)

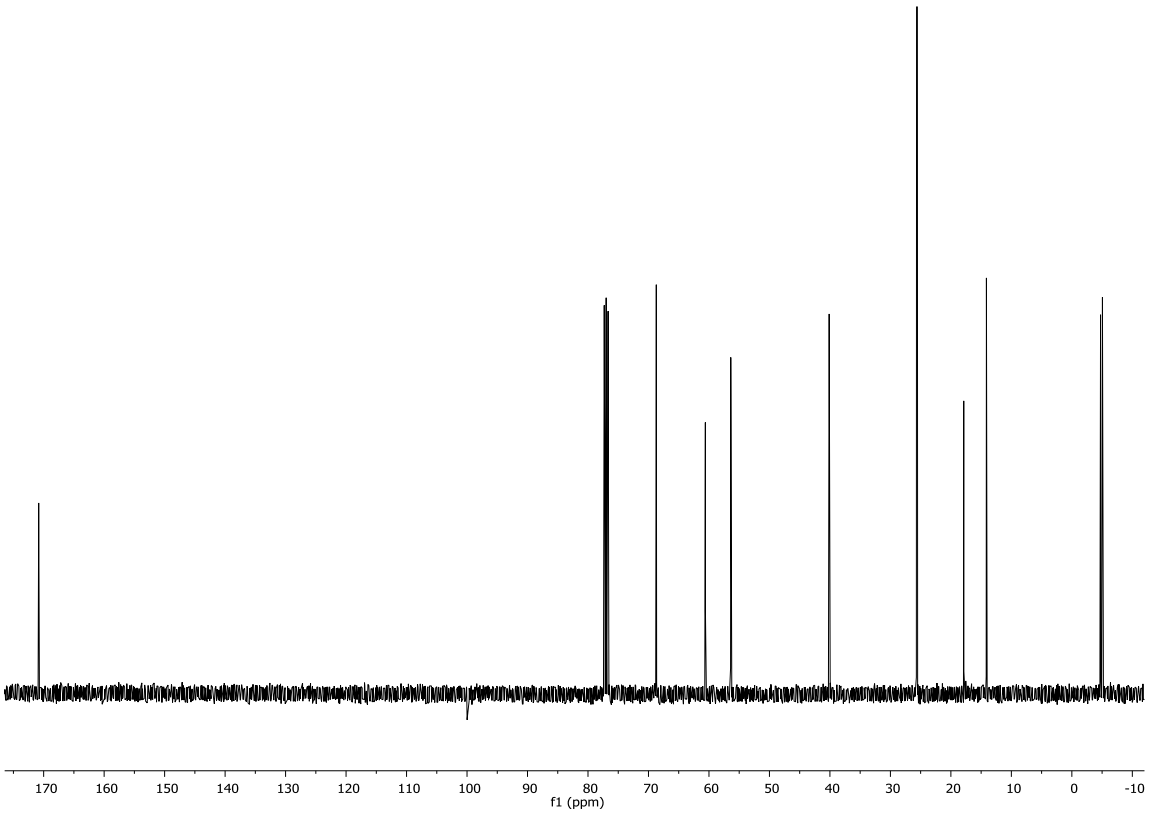

TBSO
O
OH
107
1.00
1.06
6.02
9.01
3.08
2.99
f1 (ppm)

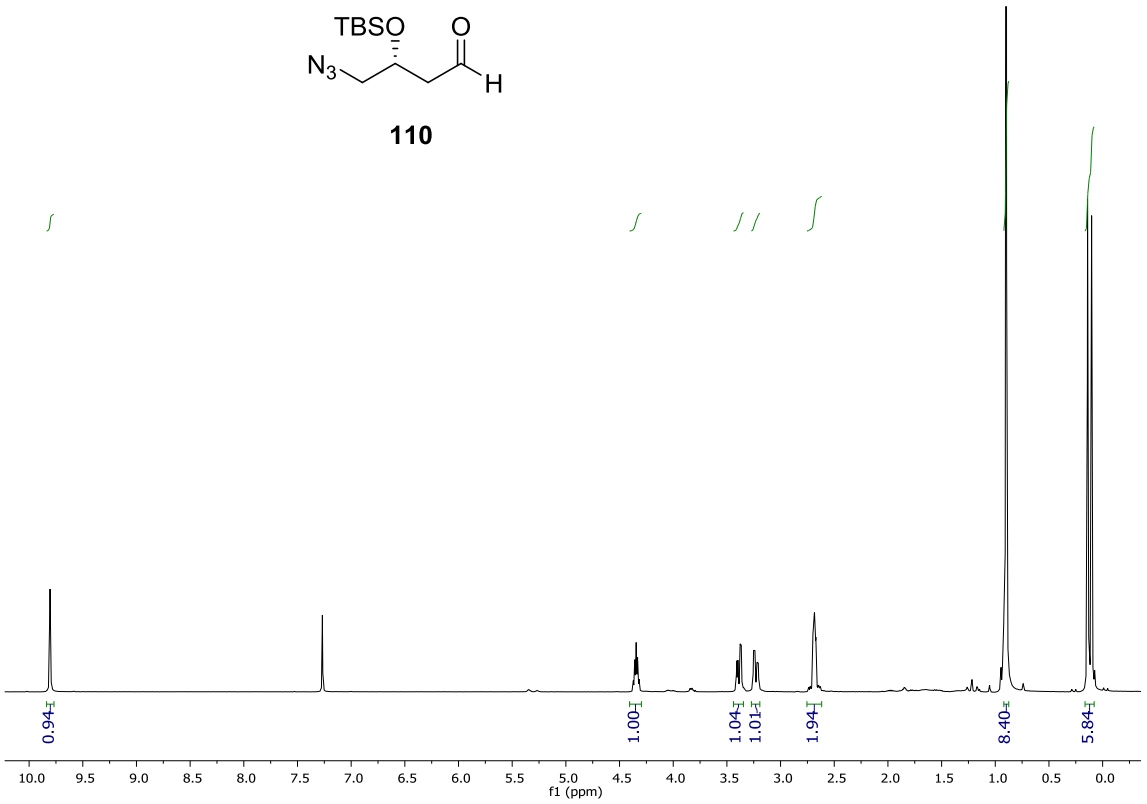

f1 (ppm)

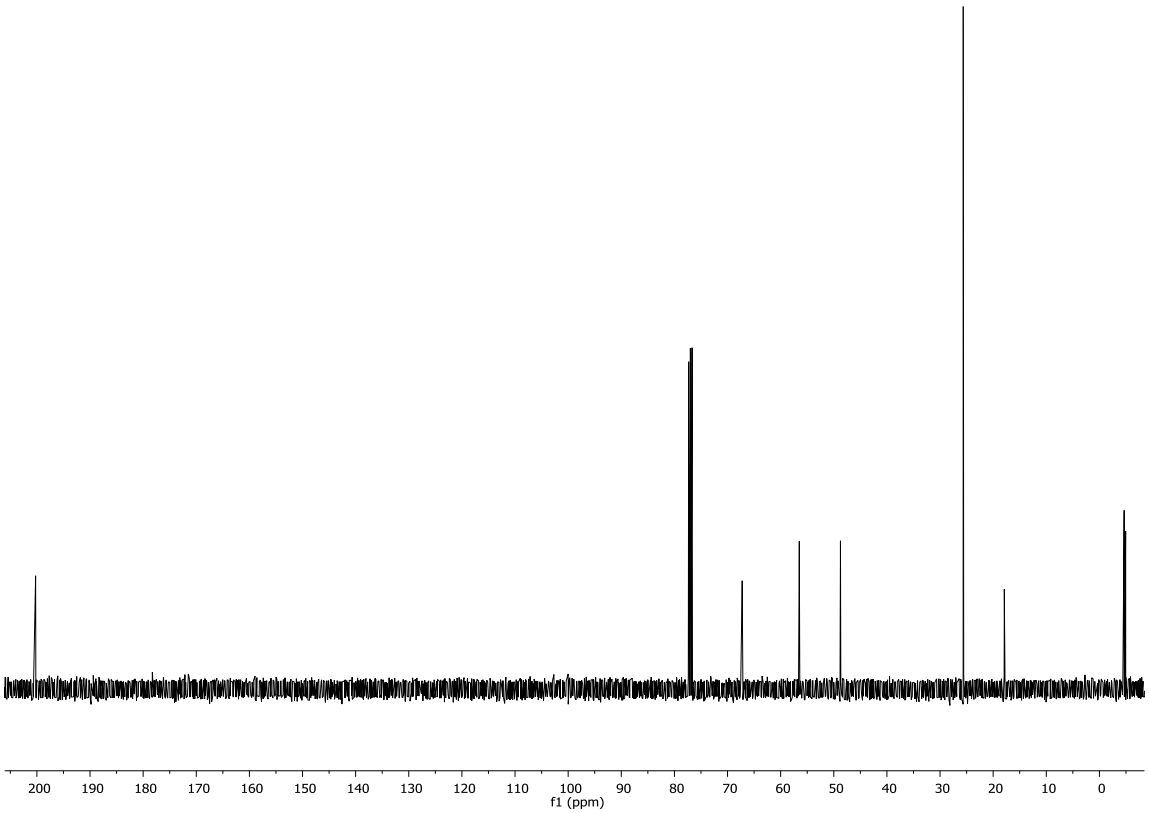

OH
O
N3
OEt
108
2.99
2.01
2.00
3.00
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

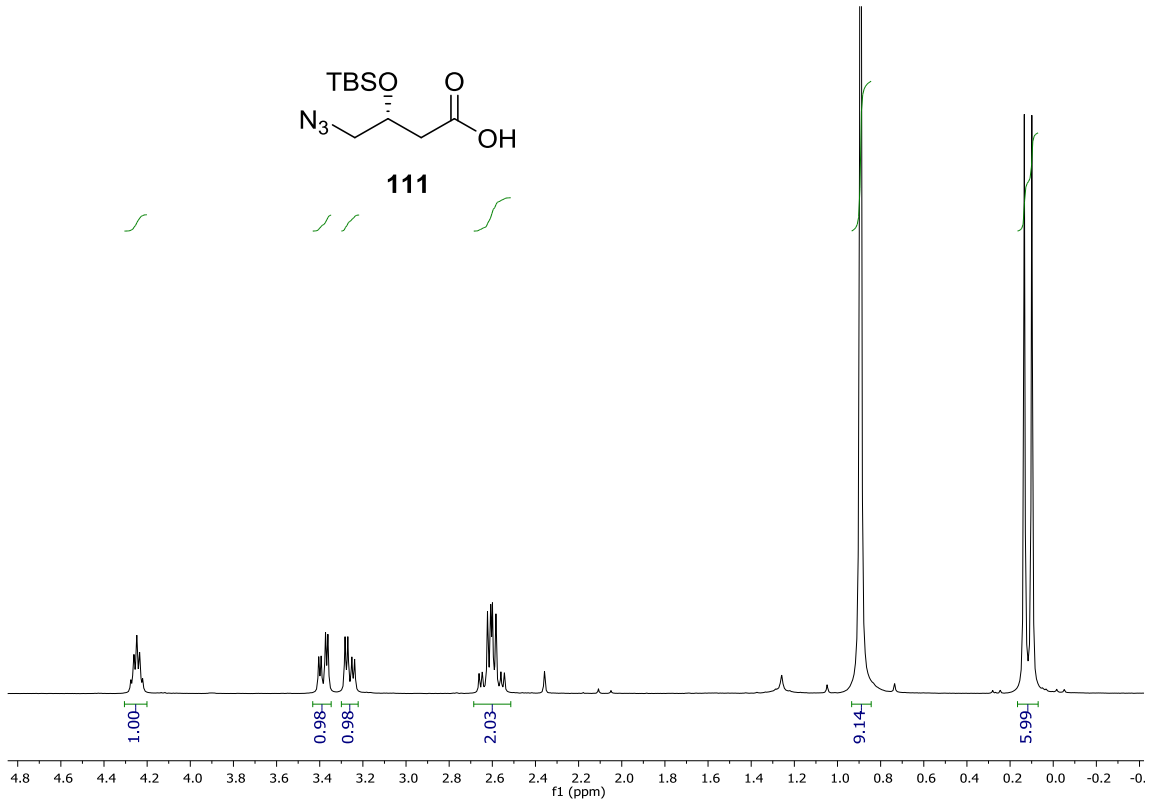

220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
f1 (ppm)

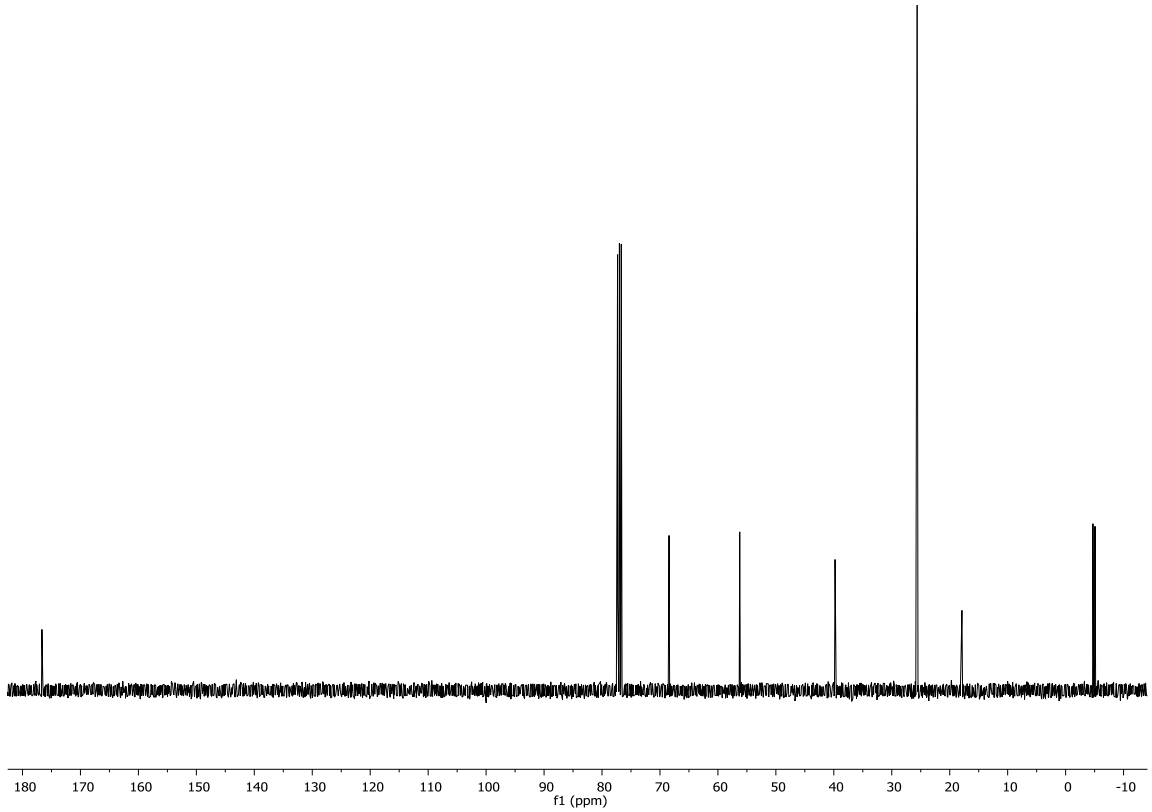

TBSO
O
N3
OEt
109
1.00
2.00
2.01
2.01
3.03
9.04
6.08
f1 (ppm)

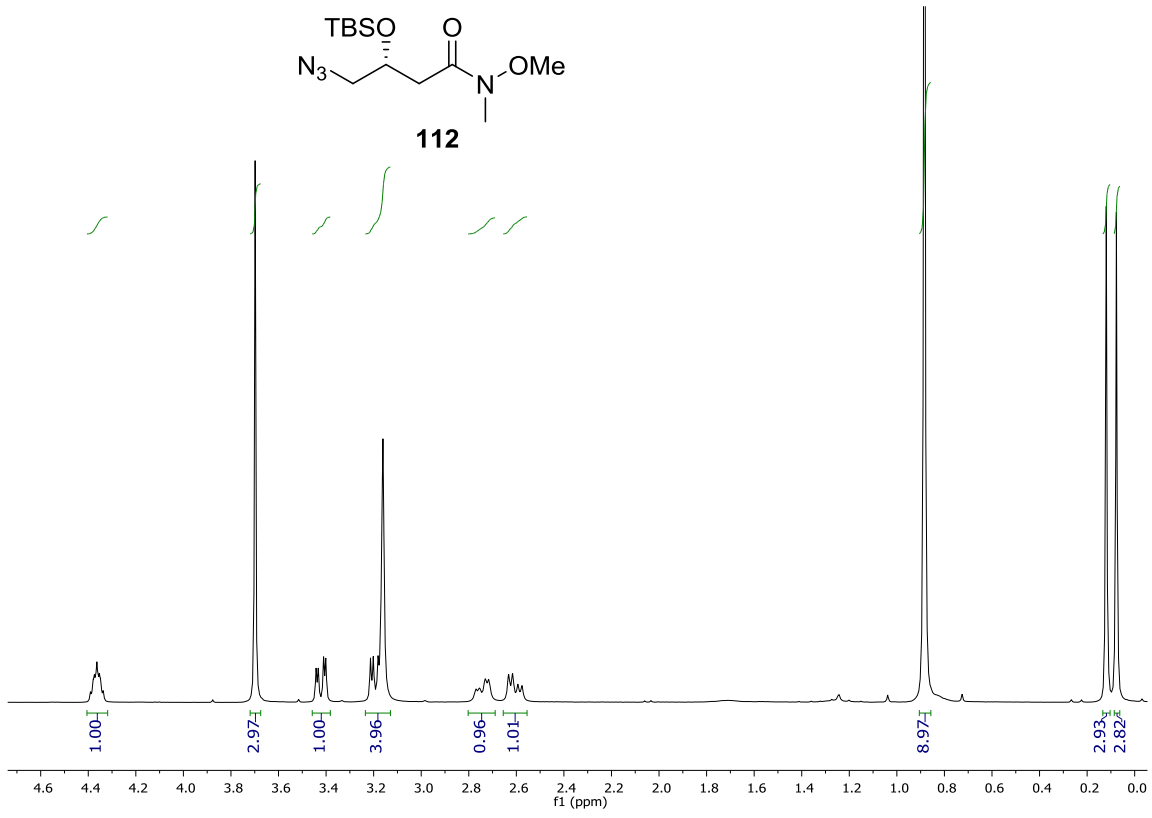

f1 (ppm)

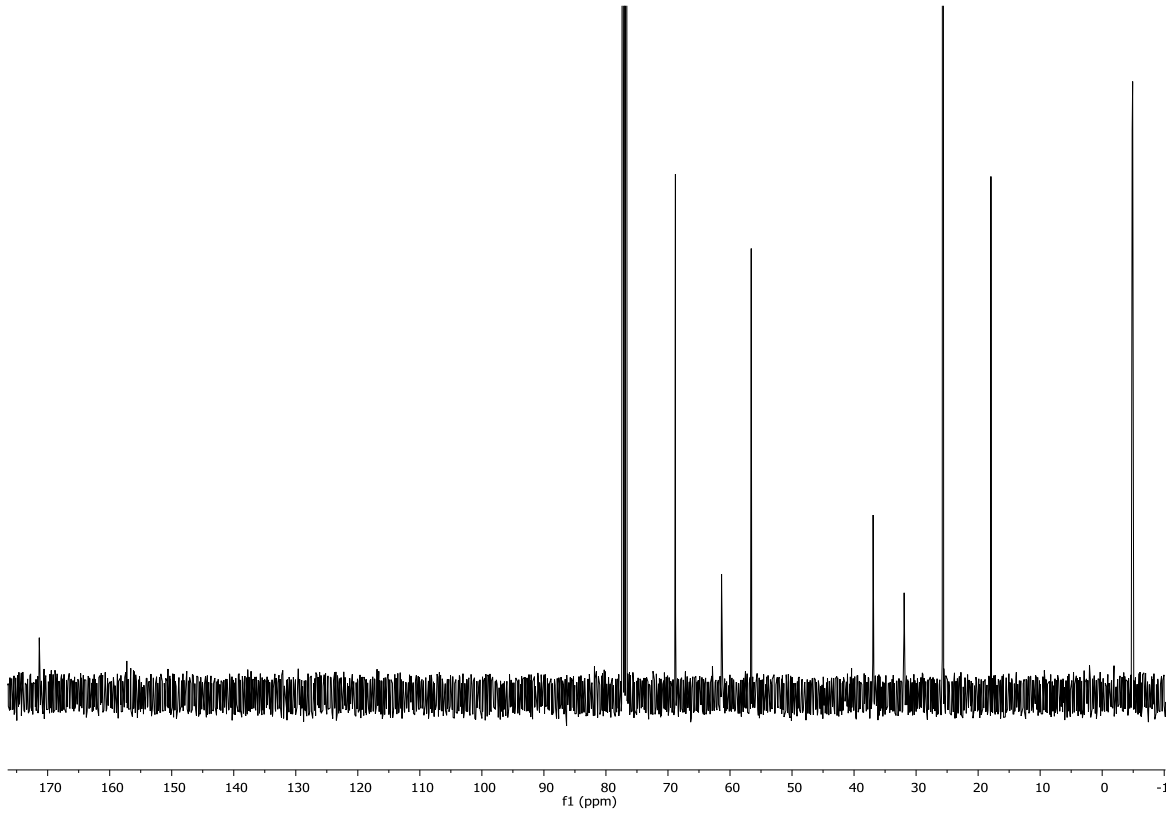

TBSO
O
N3
H
110
0.94
1.00
1.04
1.01
1.94
8.40
5.84
10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

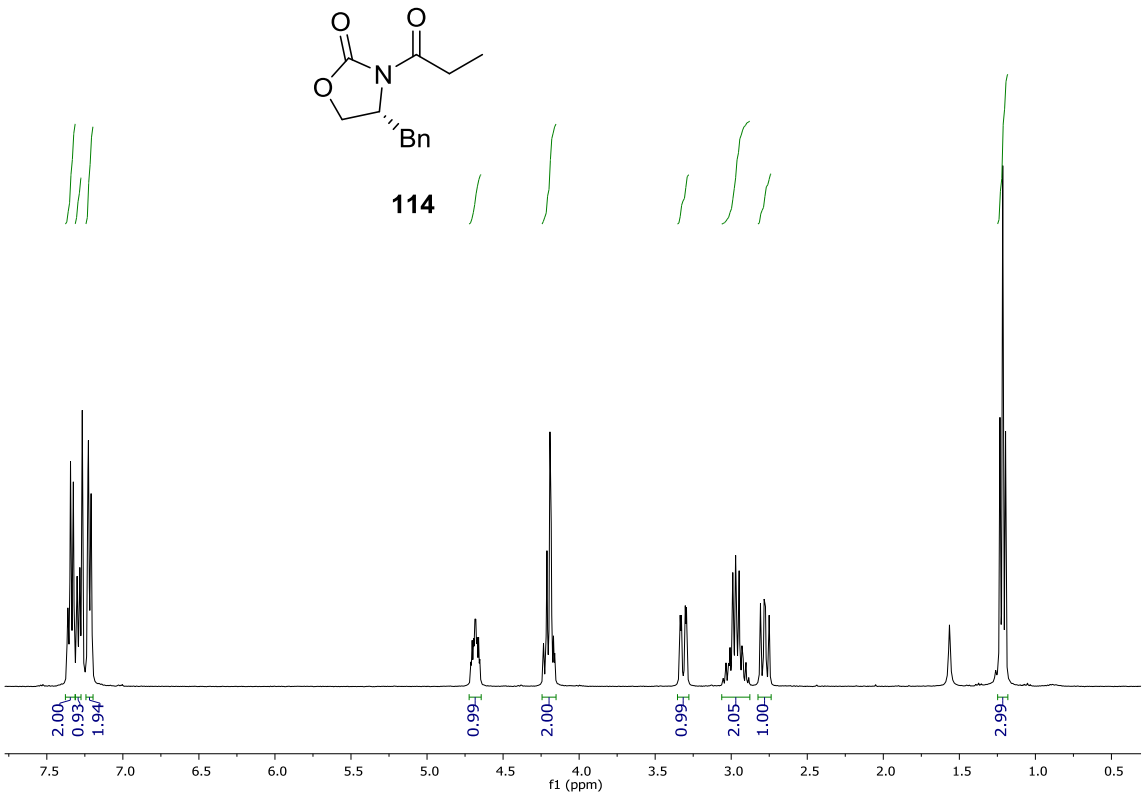

200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

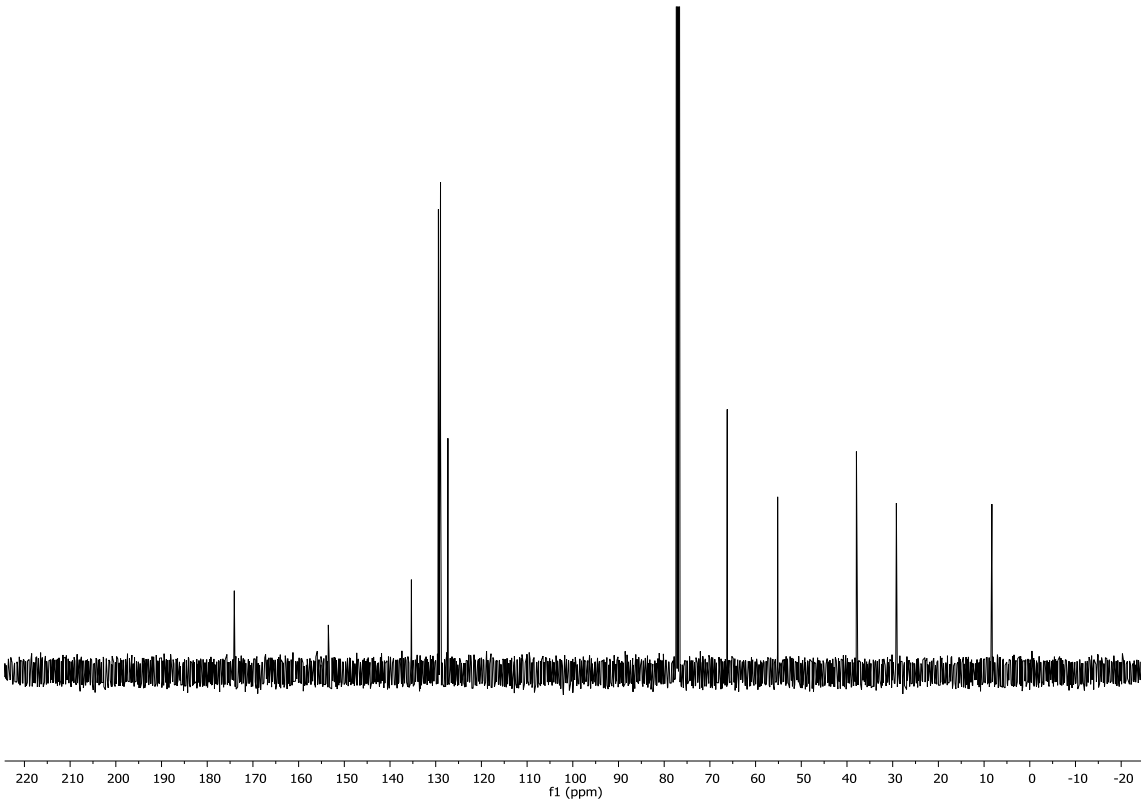

TBSO
O
N3
OH
111
1.00
0.98
0.98
2.03
9.14
5.99
f1 (ppm)

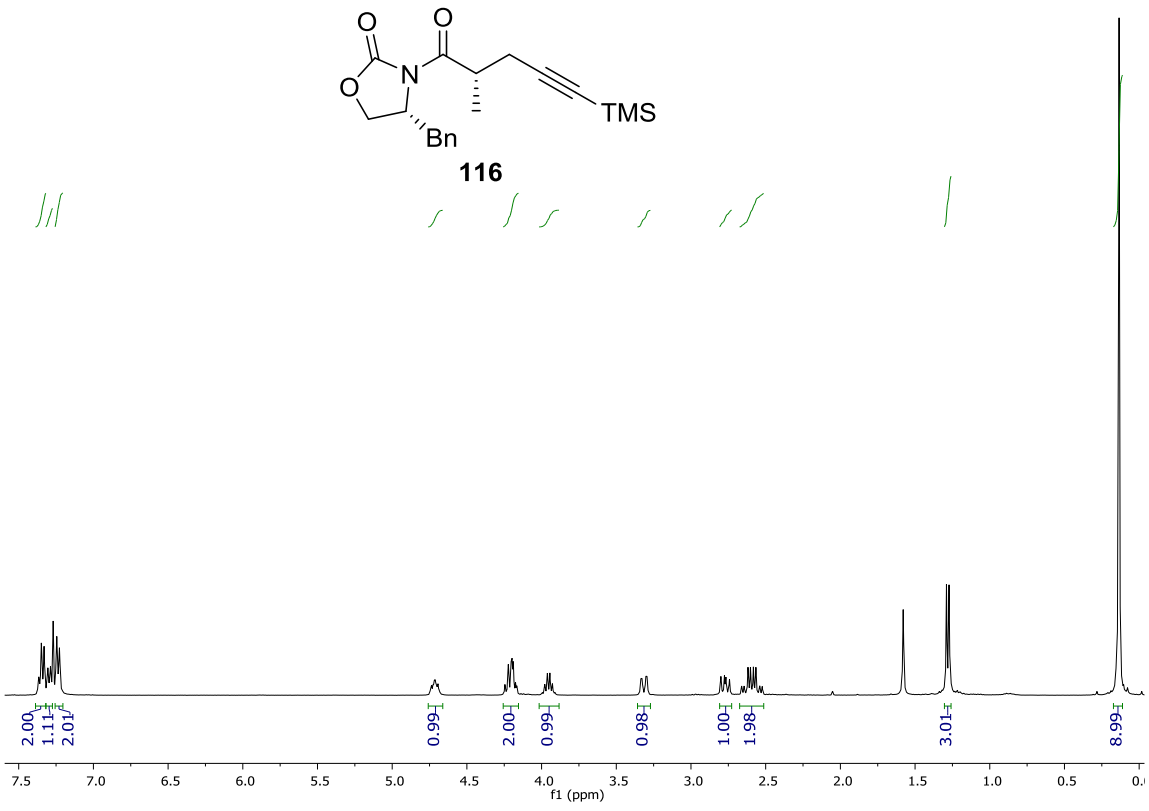

f1 (ppm)

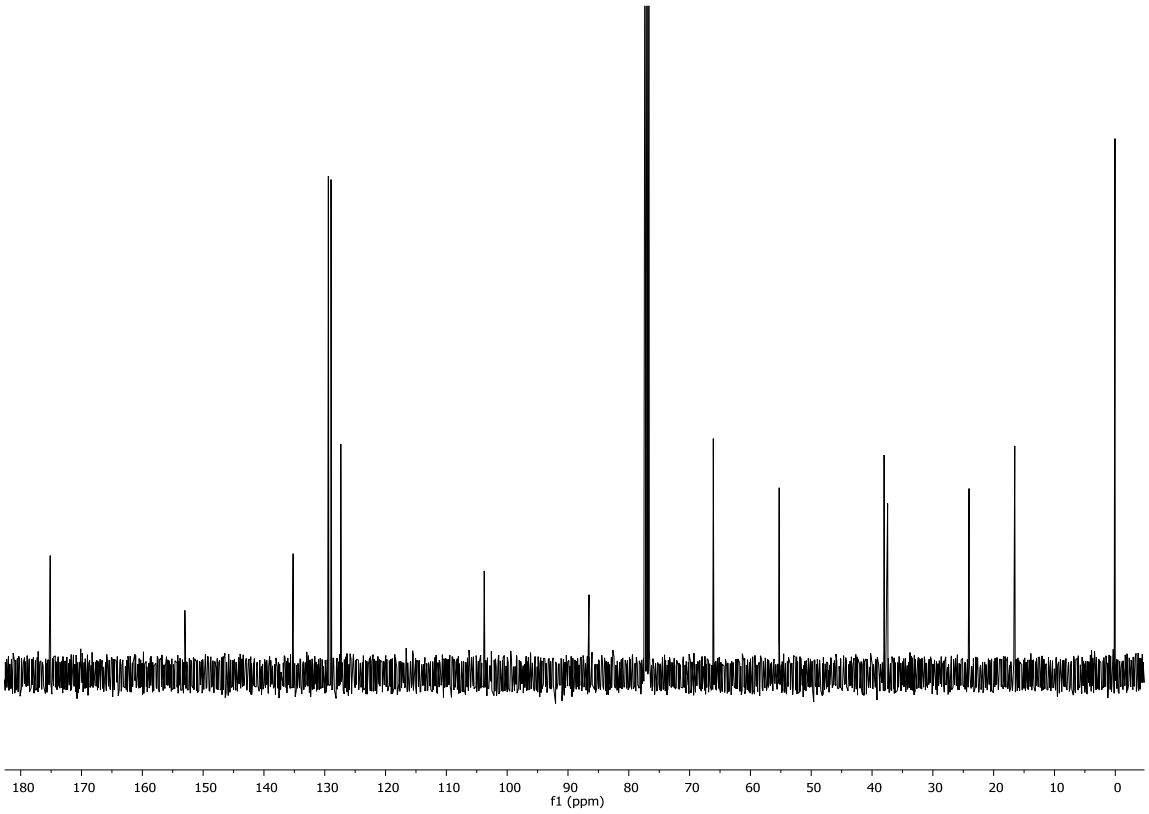

TBSO
O
N3
OMe
N
112
1.00
2.97
1.00
3.96
0.96
1.01
8.97
2.93
2.82
f1 (ppm)

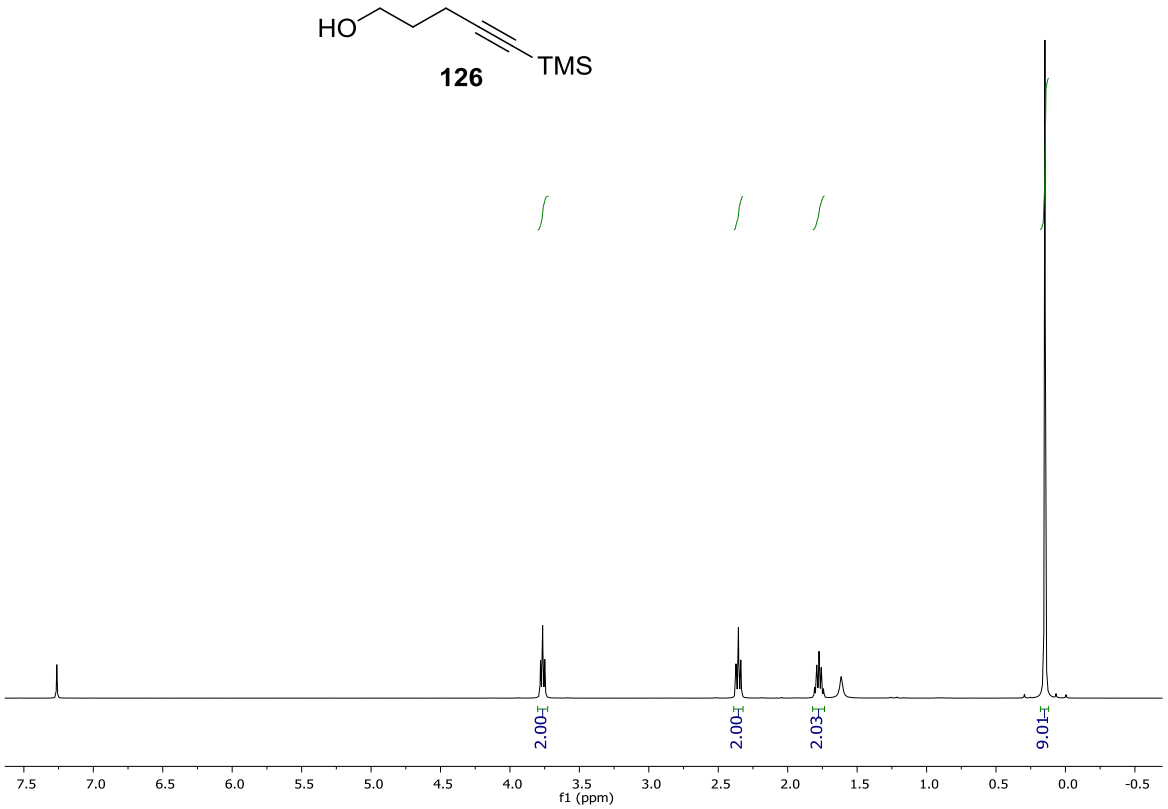

f1 (ppm)

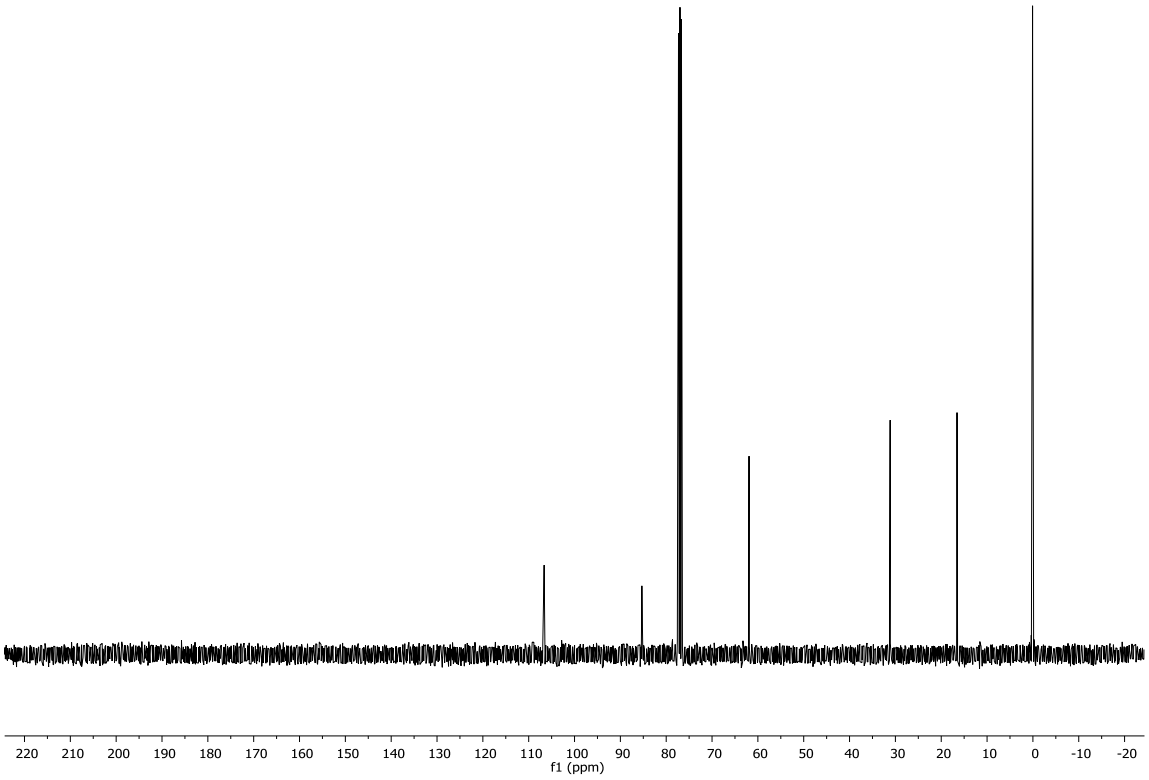

O
O
N
O
Bn
114
2.00
0.93
1.94
0.99
2.00
0.99
2.05
1.00
2.99
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

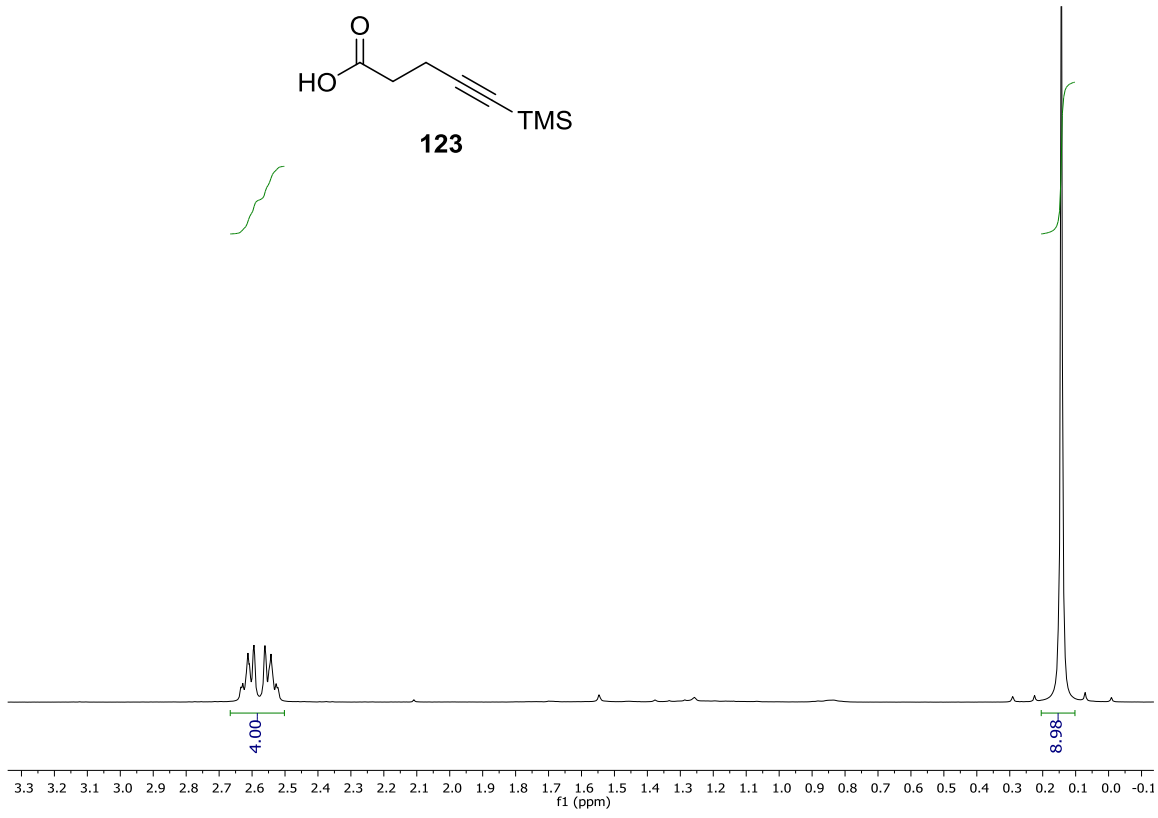

220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
f1 (ppm)

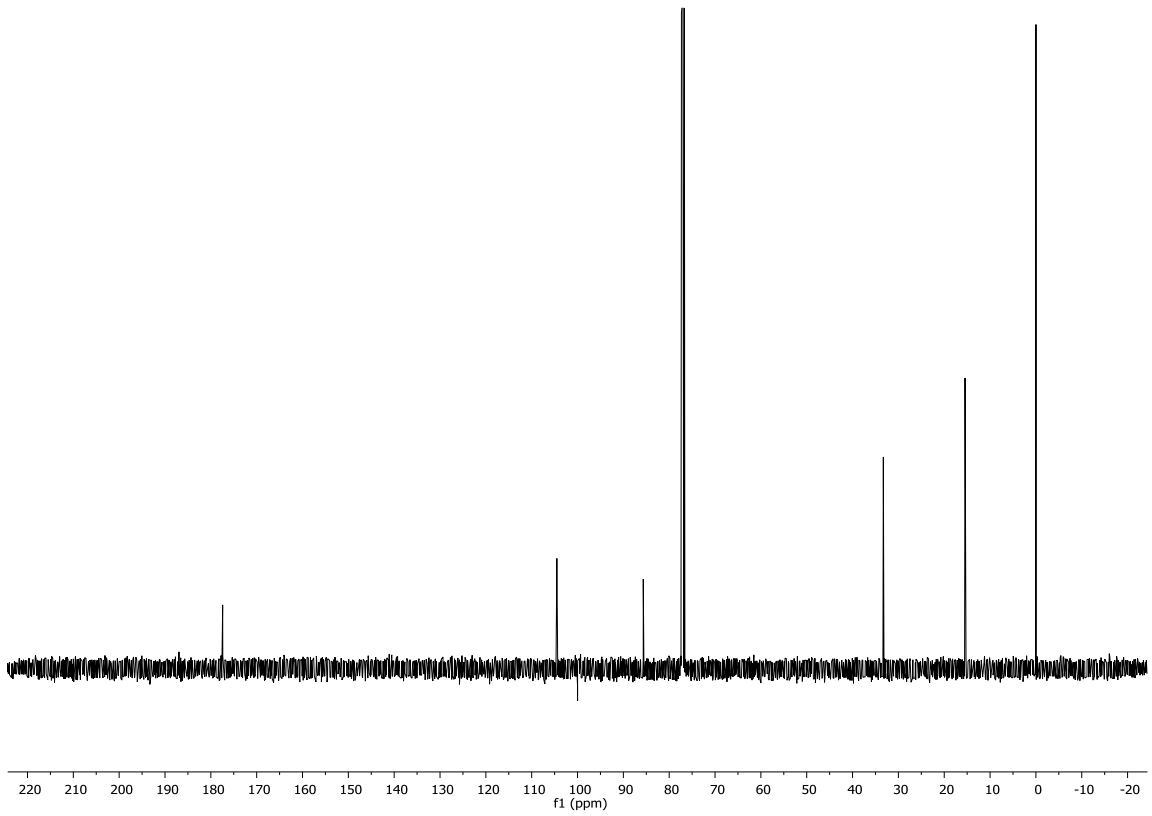

O
O
N
O
TMS
Bn
116
2.00
1.11
2.01
0.99
2.00
0.99
0.98
1.00
1.98
3.01
8.99
f1 (ppm)

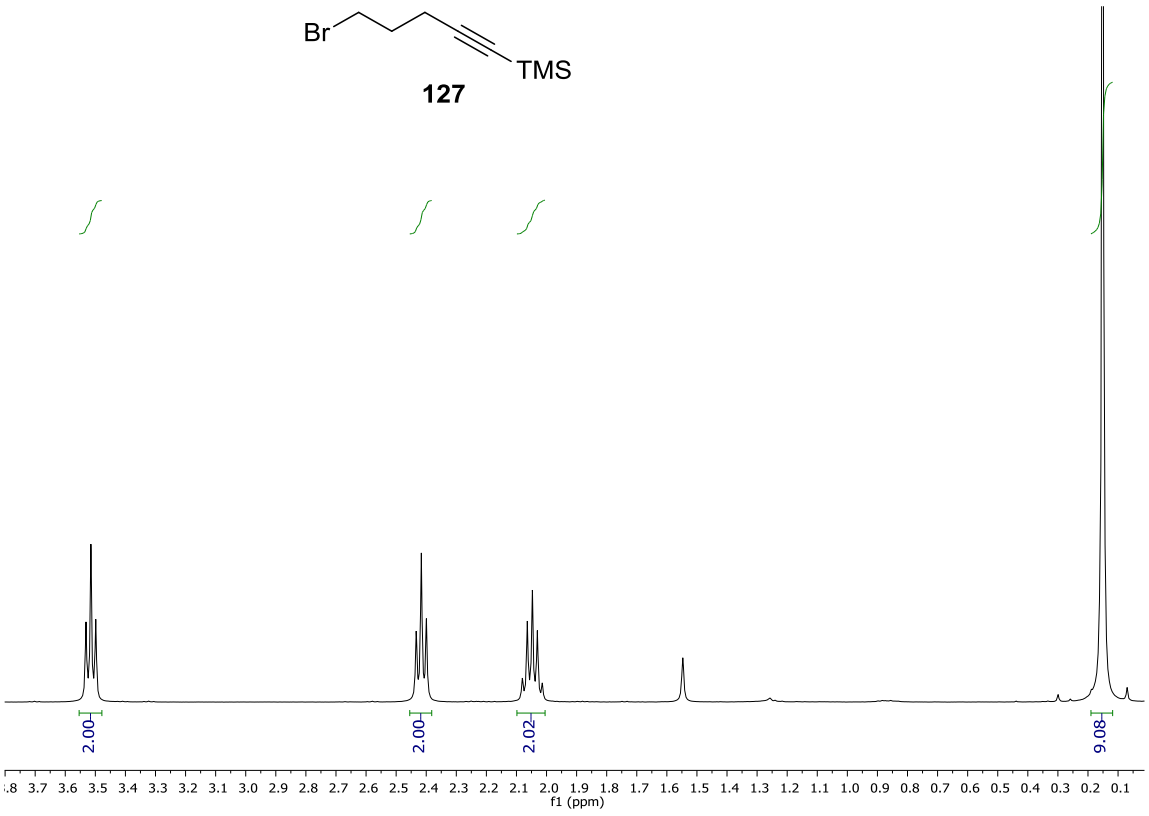

f1 (ppm)

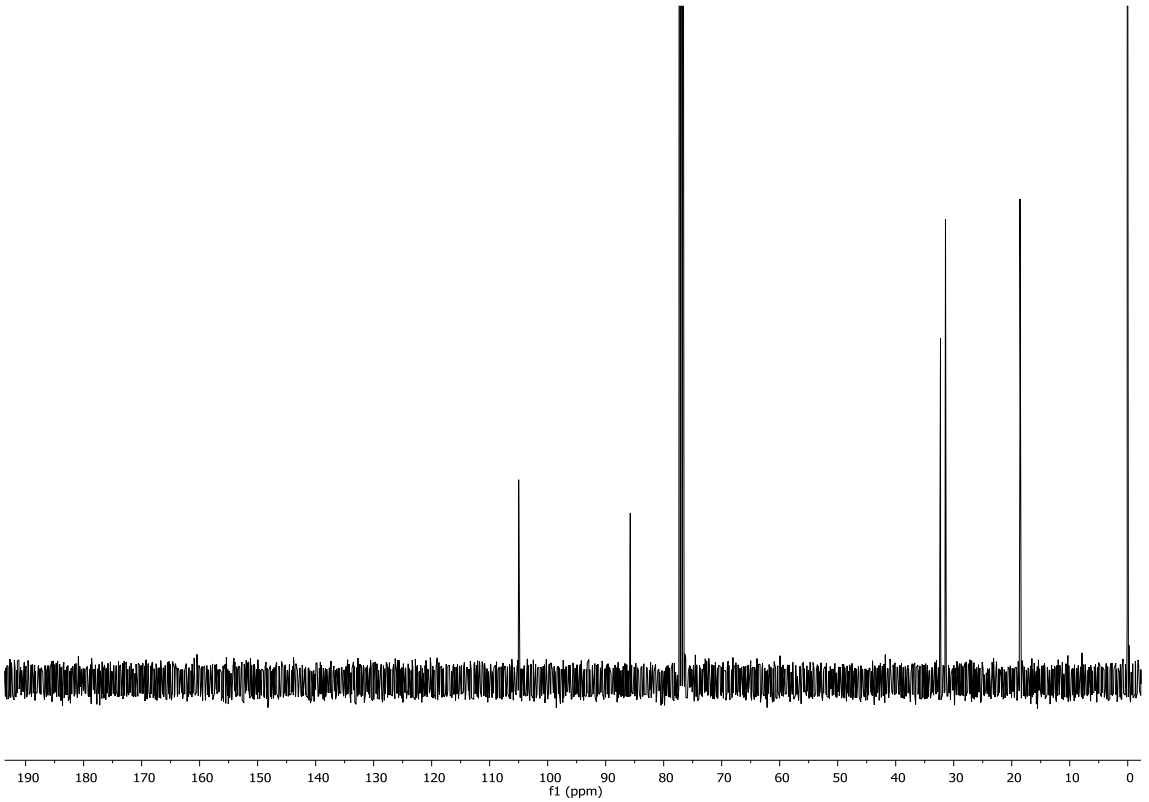

HO
TMS
126
2.00
2.00
2.03
9.01
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5
f1 (ppm)

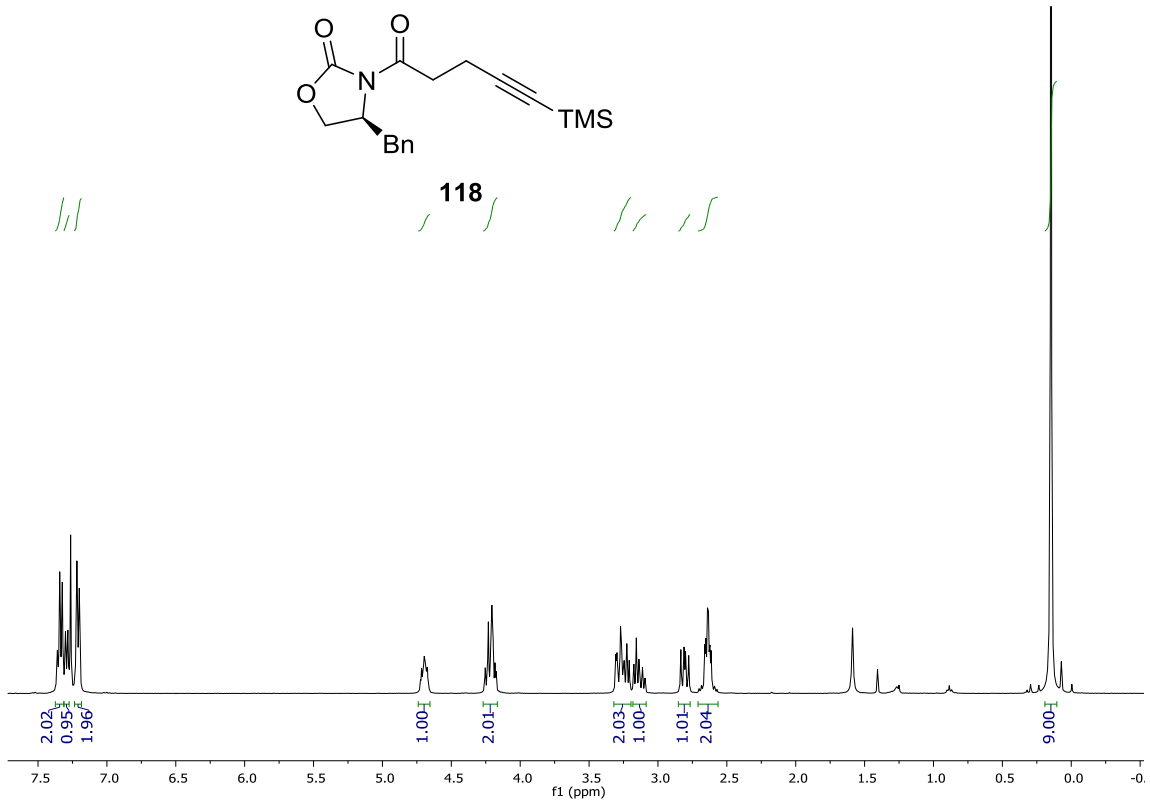

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20
f1 (ppm)

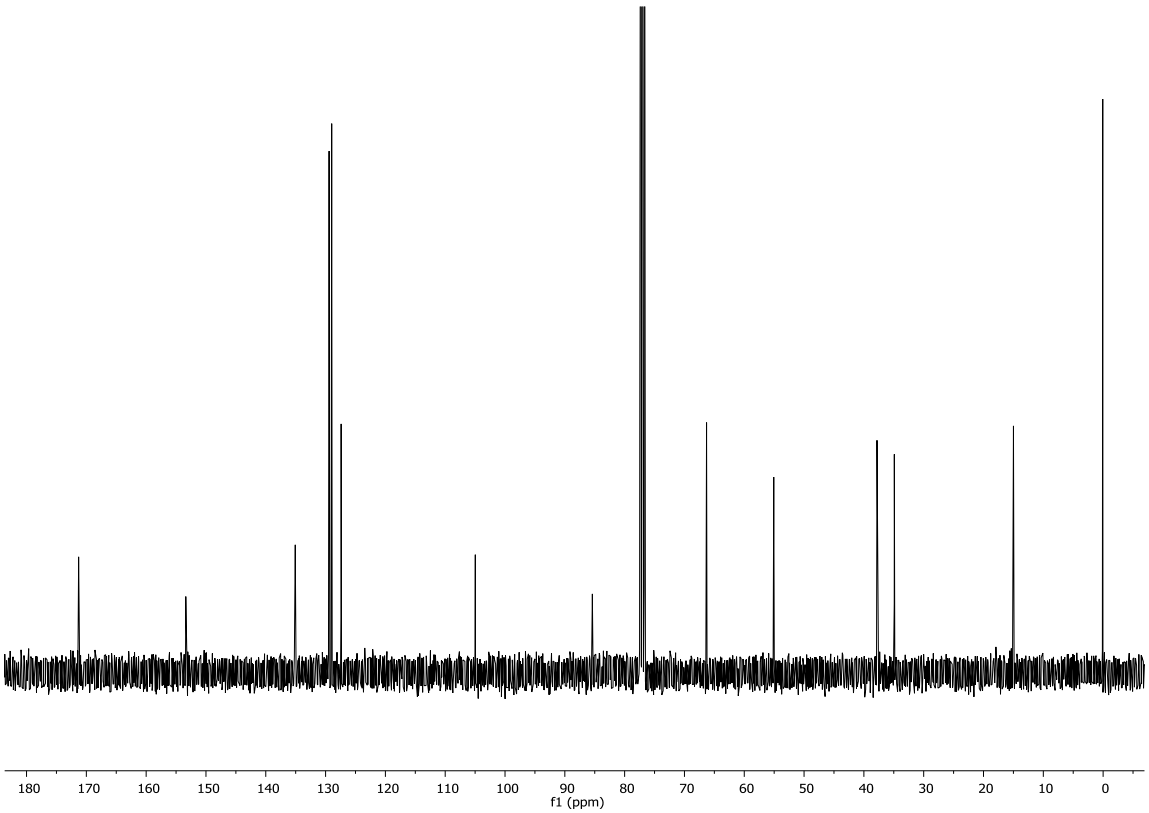

O
HO
TMS
123
4.00
8.98
3.3 3.2 3.1 3.0 2.9 2.8 2.7 2.6 2.5 2.4 2.3 2.2 2.1 2.0 1.9 1.8 1.7 1.6 1.5 1.4 1.3 1.2 1.1 1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0 -0.1
f1 (ppm)

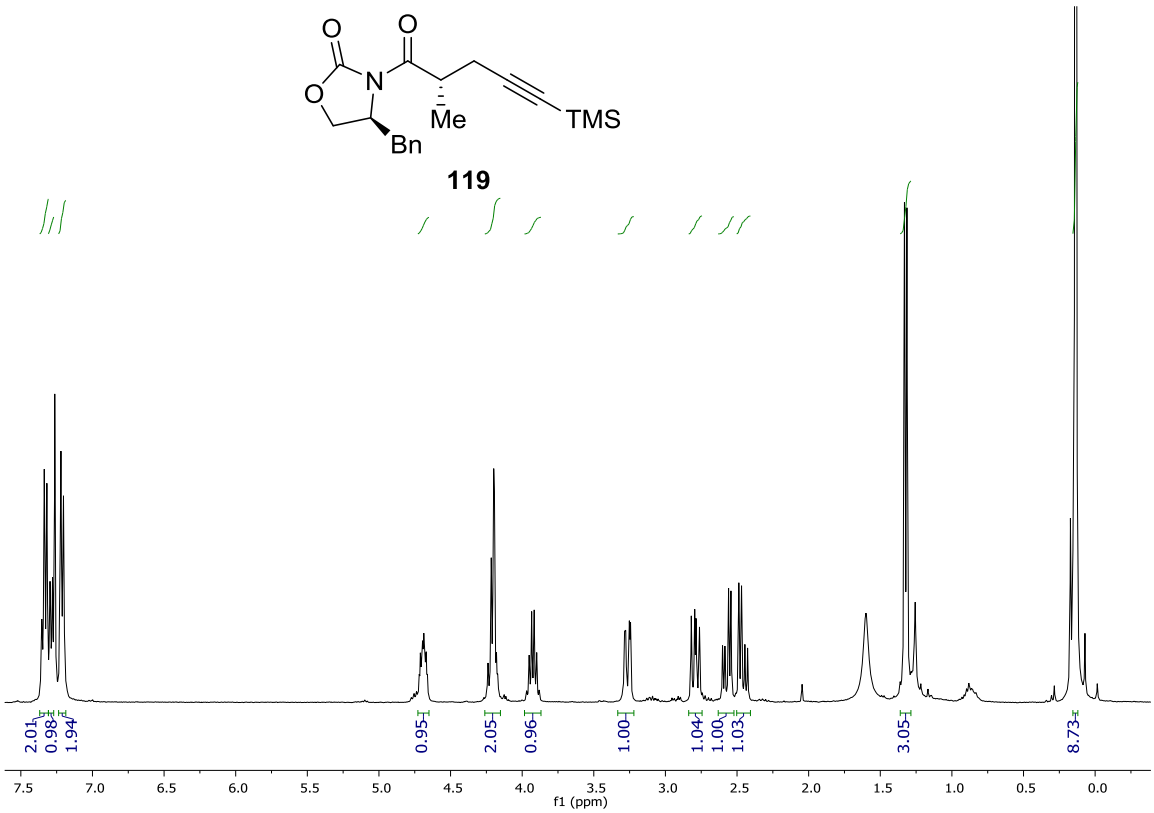

220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20
f1 (ppm)

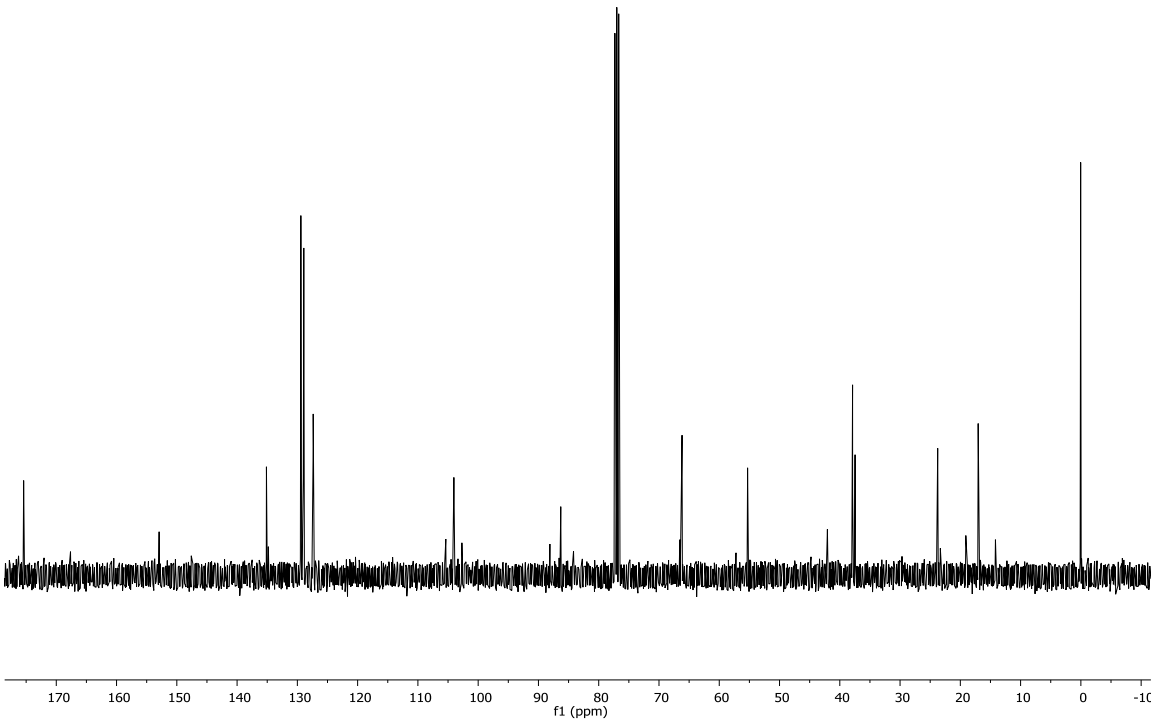

Br
TMS
127
2.00
2.00
2.02
9.08
f1 (ppm)

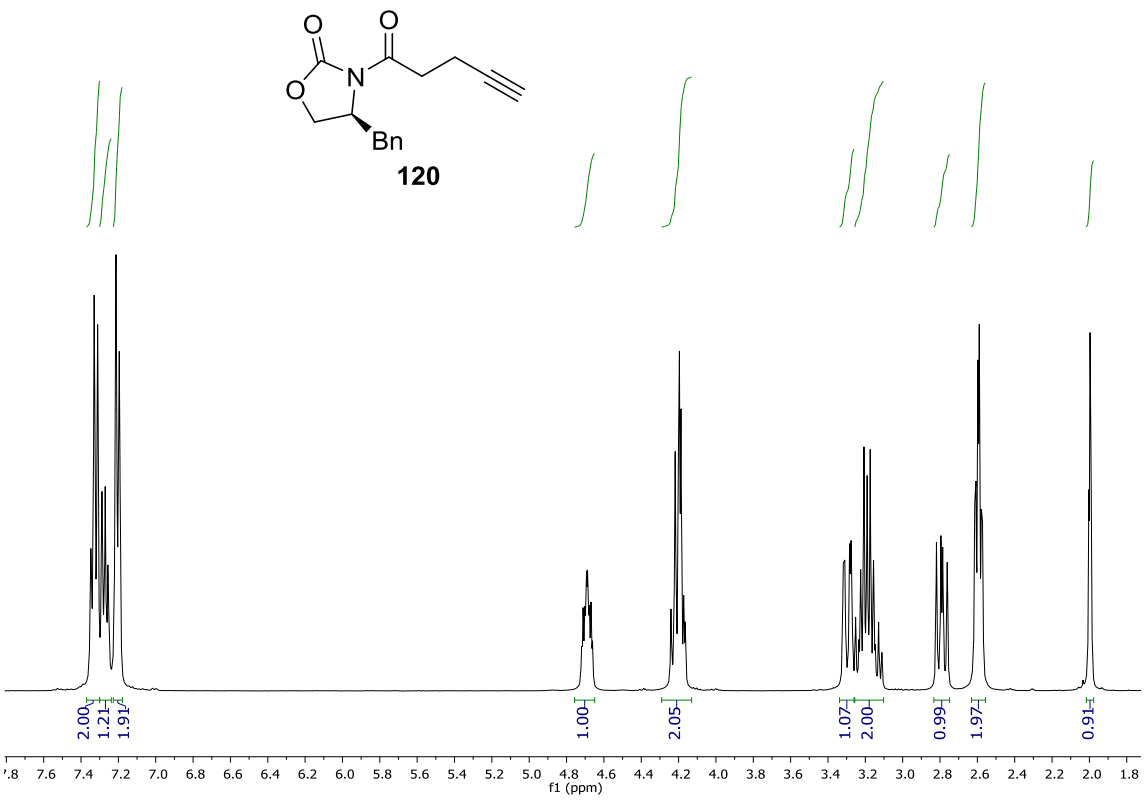

f1 (ppm)

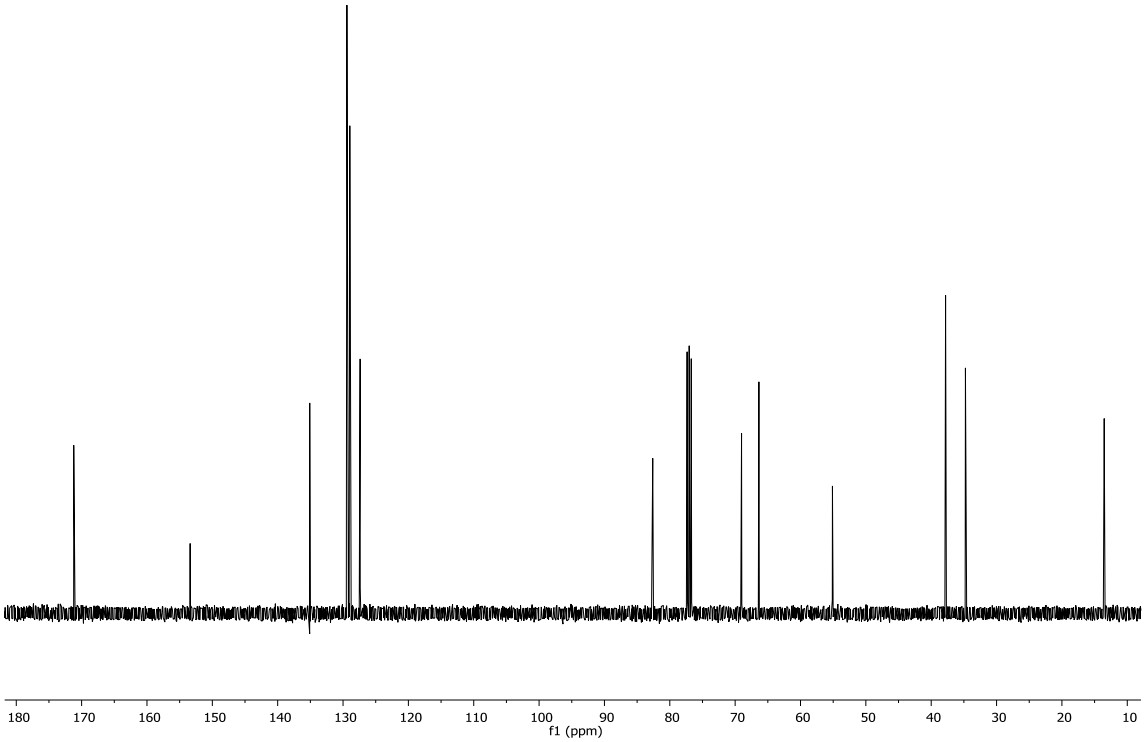

O
O
N
O
TMS
Bn
118
2.02
0.95
1.96
1.00
2.01
2.03
1.00
1.01
2.04
9.00
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

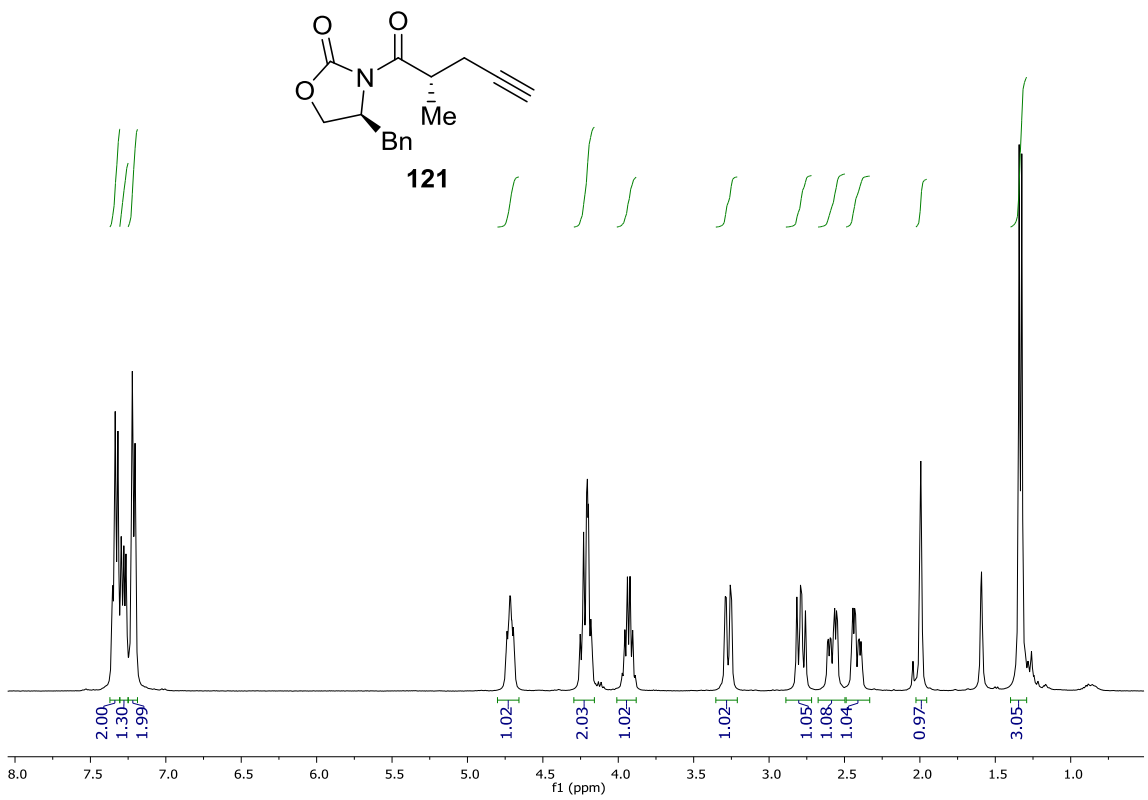

180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

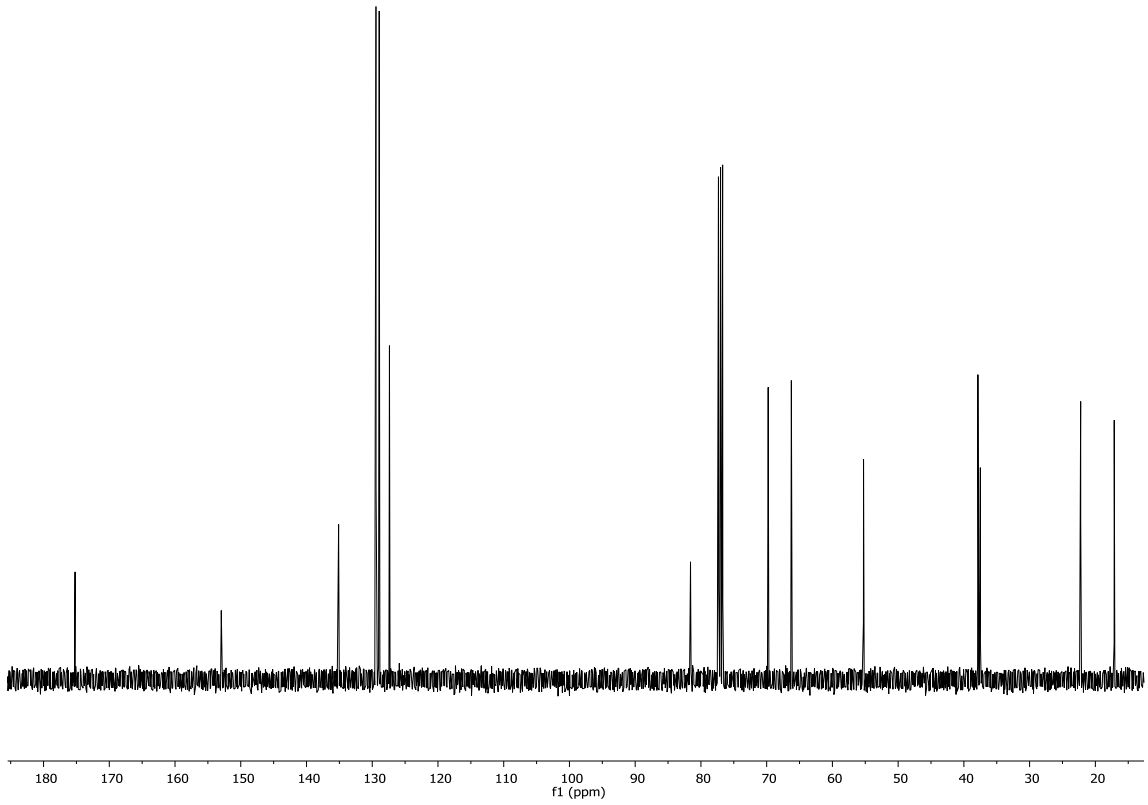

O
O
N
O
Me
TMS
Bn
119
2.01
0.98
1.94
0.95
2.05
0.96
1.00
1.04
1.00
1.03
3.05
8.73
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)


170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)


O
O
N
O
Bn
120
2.00
1.21
1.91
1.00
2.05
1.07
2.00
0.99
1.97
0.91
f1 (ppm)


f1 (ppm)


O
O
N
O
Me
Bn
121
2.00
1.30
1.99
1.02
2.03
1.02
1.02
1.05
1.08
1.04
0.97
3.05
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
f1 (ppm)


180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
f1 (ppm)

7 Appendices

7.1 Sequences

Native sequence (black) and His_6-tag/linker (red) for all proteins used in this work.

BatA

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

REGLIKLIGKTAQEILPDIEDHQFSGSDRLIDLGANSVDRAEIVMLVQEALQLSIPRVELF
GPKNIGELADLFWGKMNAS

Molecular weight 12615.39

Molecular Formula $C_{566}H_{887}N_{155}O_{166}S_3$

BatC

AHHHHHHSSGLEVLFQGP

MSLVGIEAMNVFGGTTYLDVAELARHRQLDSARFANLLMKEKAVALPYEDPVTFGVN
AAKPIIDAMSPAERDRIELLITCTESGIDFGKSMSTYIHHYLGLNRNCRLFELKQACYSGT
AGFQMAINFILSQVSPGAKALIVATDISRFLASEAGSELTEDWSFAEPSAGAGAVALLVS
DQPHVFQVDVGANGYYGYEVMDTCRPMPDSEAGDADLSLLSYLDCCENAYLEYTKR
VTNAHYRDSFQYLTFHTPFGGMVKGAHRTMMRKLLQAKTAEIEEDFQRRVMSGMTY
CQRVGNIMGATSFLSLASTIDNGQFTAPRRIGCFSYGSGCCSEFYSGVVTPDSQVRQRRF
AIKEHLNGRYRLSMAEYESIIQASNAVKFGSRNVELDKSFIGEAYESCRGKGQLFLREIK
EFHRQYEWLS

Molecular weight 48784.25

Molecular Formula $C_{2154}H_{3323}N_{597}O_{648}S_{26}$

BatD

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

SYQTIKVRFEESICFLQLDRLESANAINGQLVAECSEVLSICQEAATVVVLSGSPEAFCV
GADFGALADPATDSYGAYNGPGPLYDLWLKLATGPYVTVAHVRGKANAGGIGFVAA
CDIVLADATAQFSLSELLFGLYPACVLPFLIRRISHQRAHYLTLMTHPISAEQAFTWGLV
DALESDSEALLRRHLLRLRRLSKVAVTRYKDYMSTIGVPLAGLKSSAIAGNLEVFSDPA
NLKAIIRYVESG

Molecular weight 30399.84

Molecular Formula $C_{1372}H_{2139}N_{367}O_{396}S_9$

BatE

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

DVVQFSEVEPGIIQITMQDRENKNTFSRELVKGLINAFRHIRESERYRVVVLTGYDTYFC
SGGTKEGLLMLHEGQGKFTDMNIYSVPLECEIPVISAMQGHGIGGGFVFGLFADCVVLS
RESVYTTNFMKYGFTPGMGATYVLPEKLGLGLAEEMLLSARTYRGADLEKRGVPFPV
LPRAEVLEYALQLARDLAEKPRISLVTLKSHLVAEMRARLPTIVEQEIAMHDKTFHQPE
VKARIENVFGN

Molecular weight 31363.10

Molecular Formula $C_{1409}H_{2205}N_{381}O_{404}S_{13}$

mECH

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

ESVVRMQEIEPGIVQITMADREHKNMFSNALISGLTQAFEAIRNNETYKVVVLTGYDTY
FCCGGTKESLLDIQSGRVSFTDASVFSIPLDCKIPVIAALQGHAIGAGWALAMFSDFVVF
SQQGCYESNYMKFGFTPGAGATLIFPEKFSSGLAHEILFTGRQFTGSDLEQRGIPYPVVH
RDQVLVQAIDLARSLCQSPRIALMELKEHTANPIREKLQRTYEKEVAMHQKTFVDNPE
VLARLDALHDH

Molecular weight 31406.81

Molecular Formula $C_{1406}H_{2173}N_{381}O_{411}S_{13}$

Short mECH

AHHHHHHSSGLEVLFQGP

PRESVVRMQEIEPGIVQITMADREHKNMFSNALISGLTQAFEAIRNNETYKVVVLTGYD
TYFCCGGTKESLLDIQSGRVSFTDASVFSIPLDCKIPVIAALQGHAIGAGWALAMFSDFV
VFSQQGCYESNYMKFGFTPGAGATLIFPEKFSSGLAHEILFTGRQFTGSDLEQRGIPYPV
VHRDQVLVQAIDLARSLCQSPRIALMELKEHTANPIREKLQRTYEKEVAMHQKTFVDN

Molecular weight: 26409.22

Formula: $C_{1179}H_{1841}N_{317}O_{348}S_{12}$

BatK

AHHHHHHSSGLEVLFQGP

MMPMATITAETLGNAQFKRDYGLRYAYLGGALYKGIASKELVVAMGNSGMMAYLGT
GGMCFEEIEASIRFIQSQLSSGQPYGMNLICHLEKPDLEEQTVELFIRHGVPFIECAAFLH
MTPSLVRCRLSGLRRSAVGEIERPRRMLAKVSRPEVAAAFMRPAPEAIVRQLLEEGKVT
PYEAELSQSIPMADEVCVEADSGGHTDQGVAYALMPAMFALRDEMMAQYCYDKPIC
VGAAGGIGTPHAAAAAFTMGADFILTGSINQCTVEARTSDAVKDLLQDLNVQDTDYAP
AGDMFEIGARVQVAKRGLFFAARANKLFELYQRHNSLEEIDPKVRQRIEERYFRRSFDE
VWAETRNYYLRTEPAKVADLESNPKQKMALIFRWYFVHSTRLALSGNEDQKVDYQV
HCGPAMGAFNQWVKGTELQSWRNRYVAKIGIHLMQETARLLNARFVSMSGYLA

Molecular weight 53189.95

Molecular Formula $C_{2354}H_{3675}N_{661}O_{685}S_{31}$

Bat3 ACP3/4

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

AASRVHKTLKQTLSSVLGIDILEIGDDDAFVDLGLDSIVGVQWVRIINDRFGIDIDAVRIY
DYPSLTALVNYVTRETELAKRAQRTLVDDTGAATAQAASDKPQAAVFAHLTQSSDAS
TEPALPRRSGRAMSQVKEGLRRTLSTVLGLEPGDIADDDVFVDIGLDSIMAIQWVRVIN
SQFSLDLEAVRVYDYPTLDGLTDYVIKESELA

Molecular weight 26609.94

Molecular Formula $C_{1180}H_{1868}N_{324}O_{370}S_{3}$

Bat3 ACP3

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

AASRVHKTLKQTLSSVLGIDILEIGDDDAFVDLGLDSIVGVQWVRIINDRFGIDIDAVRIY
DYPSLTALVNYVTRETELA

Molecular weight: 12647.3

Molecular Formula: $C_{571}H_{887}N_{153}O_{170}S_1$

Bat3 ACP4

AHHHHHHSSGLEVLFQGP

SQVKEGLRRTLSTVLGLEPGDIADDDVFVDIGLDSIMAIQWVRVINSQFSLDLEAVRVY
DYPTLDGLTDYVIKESELA

Molecular weight 10697.9

Molecular Formula $C_{478}H_{742}N_{128}O_{149}S_1$

Bat3 ACP5

AHHHHHHSSGLEVLFQGP

SVADLRMELAESLAKVLYLENDDVDMDTAFSELGLDSVTGVGWIRELNQRYGTSLGA
SQLYDYPTIKVLASHLLIKLSRPVCD

Molecular weight 11200.63

Molecular Formula $C_{497}H_{776}N_{136}O_{153}S_3$

Bat3 ACP6

AHHHHHHSSGLEVLFQGP

DLSKLEAELCEMLSSTLFMPVTAADCEKPFSELGVDSIIGLEWIQAVNRRYQTSLTASM
LYQHATLKRFTLFLAVKLSRPVCD

Molecular weight 11352.06

Molecular Formula $C_{507}H_{793}N_{137}O_{147}S_6$

4M

AHHHHHHSSGLEVLFQGP

MSQVKEGLRRTLSTVLGLEPGDIADDDVFVDIGLDSIMAIQWVRVINSQFSLDLEAVRV
YDYPTLDGLTDYVIKESELATRTPVSPAKSDPEHSPIPRASAREAKQTHGDDENRRGHA
AVFESVVRMQEIEPGIVQITMADREHKNMFSNALISGLTQAFEAIRNNETYKVVVLTGY
DTYFCCGGTKESLLDIQSGRVSFTDASVFSIPLDCKIPVIAALQGHAIGAGWALAMFSDF
VVFSQQGCYESNYMKFGFTPGAGATLIFPEKFSSGLAHEILFTGRQFTGSDLEQRGIPYP
VVHRDQVLVQAIDLARSLCQSPRIALMELKEHTANPIREKLQRTYEKEVAMHQKTFVD
NPEVLARLDALHDH

Molecular weight: 42994.67

Molecular Formula: $C_{1908}H_{2978}N_{528}O_{577}S_{14}$

MacpA

AHHHHHHSSGLEVLFQGP

MNPERRNMYMEEIYTFVVSTLASSCKVQPGDIEPTTNLFADLGIDSVDFLDAVFCIEKH
YDIRIPVGQWMSAVNEGNAAMTDYFVMEHFVAQIAARAAASA

Molecular weight: 13267.96

Molecular Formula: $C_{589}H_{887}N_{159}O_{176}S_8$

MacpD

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

LNHQVMDQVFDQVEHQIAQVLGAKGGPLVAVEIDSRFSDLGLSSLDLATLISNLEAVY
GTDPFADAVAITSIVTVADLARAYAQQGVPGPSPDPLDAQLRDLRQL

Molecular weight: 14922.78

Molecular Formula: $C_{666}H_{1038}N_{182}O_{204}S_2$

MupQ

AHHHHHHSSGLEVLFQGP

MREERNWISDRLASFGDGIALIADGQQSSFNAIGQAMMSWQALLAHQNIACCTVIINGD
YSCAMVGAILALYTRHCVVVPLNRSSSERCAAVARTTGASWIIDADHSGDLQALTTQL
PTHELVAGLFDRREAGLILLSSGSSGEPKAILLSLERLFAKYRDAPRSRPRTTAAFLLLDH
IGGFNTLLHCLFSGSTLVKLDSRDAVSICRQIAEHRIQLLPATPTFLNMLLWGRAYETYD
LSSLELVTYGTETMPESTLAALNEVFPAVRFKQTYGMSELGILSTRSESSDSLWIQIATD
DVQVKVVDQILWIKTRTAMLGYLNAPSVIDQEGWLCTGDLVETRGDYFRILGRGESLI
NVGGLKVLPVEVESRLLSLPFVKDAVVWGRKSPVTGQIVAATVVLEGELDTDDTKRQI
MLHCQQGLEDFKVPRHIEFVTGRLHSDRFKKIKVASQ

Molecular weight: 51752.31

Molecular Formula: $C_{2298}H_{3659}N_{643}O_{680}S_{18}$

MupS

MAHHHHHHSSGLEVLFQGP

MTDAVSDALHTRHDSAPSAVKGTIIVSGGSQGLGLTTVRCFLEAGYNVATFSRRESPAV
TELSERADFHWQALDCTDYSALTAFVQQVEKRFGGLDGLVNNAATGVEGILSTMRVA
DIDSALDINLKGQLYLTKLVTAKLLKRGAGSVVNVSSINALRGHSGLTVYSATKAAMD
GLTRSLAKELGPRGIRVNSVSPGYFSSDMVKDLSPQTLSRIERRTPLGRLGTQQEVADLI
LYLVDRGTFVTGQNIAVDGGFTC

Molecular weight: 29454.35

Molecular Formula: $C_{1287}H_{2071}N_{375}O_{400}S_8$

TacpA

AHHHHHHSSGLEVLFQGP

MEGIYQFVQSTFSDVCNIDSEEITPTTNLFSDLGIESMDFMDVCYLIDEKYGIRIPIGEWM
GRVNEGDESAADLFVLEGFVKAVSKLVEGEVA

Molecular weight: 12378.85

Molecular Formula: $C_{551}H_{831}N_{139}O_{174}S_6$

TacpD

AHHHHHHSSGLEVLFQGP

MLNVQQIKNEVLDSIIQEVATILSEKLLPITHIENSTAFSELSLTSLDLAELISNLEARYEV
DPFEELVAITSIVTIEDLASAYALSLSGNTEDSHDLLSDELKAIKNQVR

Molecular weight: 14266.01

Molecular Formula: $C_{631}H_{1010}N_{166}O_{207}S_1$

TmlQ

AHHHHHHSSGLEVLFQGP

NRTWISDRVNSFGDSLAVVEGDNQTTFNAITKSTASWLENESIAPQSVIVINGIFNATSIG
LILCAYIKECIIVPLFNYDEEKVRTILDMVQPDLLCDTRSDTARCILFDTPPNSKPALFREI
NTRRESGLVLYSSGSTGTPKAILLSLDKLLHRYQQHNQRSALTIAGFLFFDHIGGFDVM
MQCLLTGNTLVSMTQRTPDDVCKAIEKHRINVLPTTPTFLNMLLINRAYQRCDLSSLTV
IAYGSEVMPKATLNLLHQALPEVMLKQTYGMSELGVLPTESKMGNSLWLKIKKAKYK
VQDGVLWIKSPTAMLGYLNKESDELNDEWLCTGDLVEEQGEYIRILGRQSTVINVAGE
KVFPAEIEALLLQIPYVKNSLVWGKKSHITGKIVAATIFIDEDIDQKQAKKHISDFCKQA
LEPYKIPRYFEFVNDPYHSERFKKINKI

Molecular weight: 52091.05

Molecular Formula: $C_{2339}H_{3705}N_{623}O_{683}S_{19}$

TmlS

AHHHHHHSSGLEVLFQGP

MESENGTVIISGGSRGLGFEIASQFLAQGYAVATFSRGSSEQVVALAEEPRFFWKSVDG
SDYQALTEFLKEAQKKLGNIVGLVNNAAIGADGVLSTMRTSDIDRAIDVNLKAQIYLSK
LVSKKLLQNRDGFIINISSIMGVRGLPGVSIYSATKAAMDGITRSLAKELGRKGIRVNSV
SPGYFSSDMVKDLSDDILRKIERRTPLGRLGTQDEIAKLVLFLATEGKFITGQNIVIDGGF
TC

Molecular weight: 27959.94

Molecular Formula: $C_{1239}H_{1996}N_{348}O_{375}S_6$

7.2 Sequence Alignments

Full sequence alignment of ECH_2 domains presented in Figure 63.

```
BatE                        ---MNRSDVVQFSEVEPGIIQITMQDRENKNTFSRELVKGLINAFRHIRESERYRVVVLT
MupK                        -----MDSVIDFQARDDGVVVLTLQDREHKNTFTPALIDGLIQAYARINASDAYKAVVIT
CalR                        --------MVDLSEIEPGIVLLQMQDRIHKNAFTDDMTLALLNAFDTIAHHDRYKVVVLT
BaeI                        ----MPNAAVELLEIEPDIIQVTMQDKAHKNAFSKELTDGLMEAFTRIKKSSVYKAVILT
VirE                        --------MVH-LSWQGPVAVIRMADHEGRNTFGPALSRGLVDAVDRAVAEERTRVLVVE
pksI                        ----MTHSVVELIEIESAIIQVKMQDRTHKNAFSQELTDDLIQAFEYIRQNPKYKAVILT
TmpE                        MNKTIINSAVYTVENANGVVQITLEDRANKNMFTRDIIAGLMQAYKNIDENQSCKVVIIT
TaY                         -----MAPVVTLHEVEEGVAQITLVDRENKNMFSEQLVRELITVFGKVNGNERYRAVVLT
OocC                        ----MSTDVIDFQEIEPGIAHIRMQDRAYKNTFSRALVDALILAFKQIESAERYKAVILT
CurF                        -----NSEVVQLTELGNGVVQITMKDESSRNGFSPSIVEGLRHCFSVVAQNQQYKVVILT
JamJ_ECH_1_14-255           -----NSEVVQLAELGDGIVQITMNDEKNRNSFSKNFIDKLFRCFELVNQNPIYKVVIIA
Bat3_ECH_1_2986-3226        ------ESVVRMQEIEPGIVQITMADREHKNMFSNALISGLTQAFEAIRNNETYKVVVLT
PedI_ECH2_2878-3125         ------SEVARLSRYRNGVLLMTLCERQSKNTFSRIFVEGVIGAFEHISGNPSYKVVVLT
OnnB_ECH_2_2961-3192        ----------------GVVLVSLCDRENKNMFSKAFLQGFDEAFEHIRGNAAYKVVIIT
NspA_ECH_2_3211-3455        -------DVVHVHAFGNGVVQVTLADTANKNMFSNAFSEGLTQAFEHIESVPTYKVVVLI
DipP_ECH_1_2970-3215        ------SHVIKLEKYYNDILILTICDRENKNMFSSSVVKGITDAFSYINSSVQYKVIIIT
PsyA_ECH_1_1536-1762        -------------------------DRASKNSFSEAFLDGMCEAFDHIRAQEAYKVVVLA
PhmE_ECH_2_2206-2441        ----------------GVLLVTLCDRNSRNTFSKAFIAGINEVFALVRDNPRYKVLVLT
PhmI_ECH_2_1102-1349        ------SKVVKLEIFSDGVAVLTMDDRDSKNTFTEALAQGLTEAFAWVEKTFSIKVVVLT
Fr9GH_ECH_2_3120-3369       ----LPHTCIELTLDDLGVAVARMCDPDTKNAFTPAFVAAMEAVMTWAGETPACKVLILT
OocJ_mECH_2_2220-2467       ----LTSAAIEAEVYDNGVLLVRLCERDSRNKFSKAFVAGFEEIFAHIRQNQRYKVVVLT
                                                :   :* *   .   .            :.:::

BatE                        GYDTYFCSGGTKEGLLMLHEGQGKFTDM-NIYSVPLECEIPVISAMQGHGIGGGFVFGLF
MupK                        GYDSFFCSGGTREGLLALNQTRGDFTDA-NVYSLAFECKIPVIAAMQGHAIGGGLVMGLF
CalR                        GFDSYFSTGGDQASLLRLHEGRGHFATT-NLYSLALNCRIPVISAMQGHGIGGGFVIGLF
BaeI                        GYGHYFASGGTQEGLLRIQQGITKFTED-RMYSLALECDIPVISAMQGHGIGGGFVMGLF
VirE                        GMPELFCAGGSQRELVHFARGDGSFDTD-DFFRVLARCPLPVVAAVQGHAIGGGLVLALY
pksI                        GYDNYFASGGTQEGLLRIQQGLTKFTDD-NLYSLALDCEIPVIAAMQGHGIGGGFVMGLF
TmpE                        GFDSYFCSGGTQDSLLSLNDTQGNFSDV-NIYSLPLECRVPVISAIQGHAIGGGLVMGLF
TaY                         GYDTYFALGGTKAGLLSICDGIGSFNVT-NFYSLALECDIPVISAMQGHGVGGGFAMGLF
OocC                        GYDNYFATGGTQEDLFALQAGKGKFTDT-NIYSLPLECPIPVISAMQGHAVGGGLVMGLF
CurF                        GYGNYFSSGASKEYLIRKTRGEVEV---LDLSGLILDCEIPIIAAMQGHSFGGGLLLGLY
JamJ_ECH_1_14-255           GTSHYFSTGATQEQLQSIFREEIKFTDFYEILTLALDCPLPVIAAMEGHALGGGLNLGLY
Bat3_ECH_1_2986-3226        GYDTYFCCGGTKESLLDIQSGRVSFTDA-SVFSIPLDCKIPVIAALQGHAIGAGWALAMF
PedI_ECH2_2878-3125         GYDNYFACGGTKQGLLDIQSGKARFTDE-QSYRMPLSCDIPVIAAMQGHAIGAGWAMGLF
OnnB_ECH_2_2961-3192        GYDSYFACGGTKESLQAIQQGAEKFTDT-RIYGRPLECEIPVIAAIQGHALGGGWSMGMY
NspA_ECH_2_3211-3455        GYNQYFACGGTKSGLQDIYNGRARFTDN-DIHSLPLRCHLPAIAAMQGHAIGGGWSLGMF
DipP_ECH_1_2970-3215        GYENYFACGGTKEGLLKIQKGISRFTDE-KSYCMPLYCKIPVIAAMQGHAIGAGWAMGLF
PsyA_ECH_1_1536-1762        GFDHYFCCGGTKEGLLSIQQGQVQFTDF-KVFSLPLQCELPVIAAMQGHAIGAGWSLGLF
PhmE_ECH_2_2206-2441        GYDSYFACGGTQETLLSIQGGTARFTDD-KIFSLPLECEIPVIAAMQGHGIGPGWALGMF
PhmI_ECH_2_1102-1349        GYGHYFACGGTREALLAIQAGRARFTDS-RVTSLPLECNIPVIAAMQGHGIGAGWALGMF
Fr9GH_ECH_2_3120-3369       GYGHYFATGGTRQGMQAIQQGNATFTDA-RLYELPLACPVPVIAAMQGHAIGAGWSMGMV
OocJ_mECH_2_2220-2467       GYDHYFASGGTMETLQAIQRGEAKFTDE-KIFALPLECDIPVIAAMQGHGIGPGWAVGMF
                            *    *. *.    :          .             * :* ::*::**..* *  ..:

BatE                        ADCVVLSRESVYTTNFMKYGFTPGMGATYVLPEKLGLGLAEEMLLSARTYRGADLEKRGV
MupK                        CDIVVLSRESLYTANFMKYGFTPGMGATYILPSRLG-SLGQEMLLAAQSYSGALLEARGA
CalR                        ADVVILSRESVYTTNFMRYGFTPGMGATYIVPLKLGTALGHEMLFLARNYRGEELKQRGA
BaeI                        SDFVILSKESVYTANFMKYGFTPGMGATFIVPEKLGASLGHEMLLNGGHYRGGDLEKRGV
VirE                        ADLAVFSERSVYAANFMRYGFTPGMGATHLLPARFGHQLGTEMLYTARNHRGAELRERGA
pksI                        ADIVILSRESVYTANFMKYGFTPGMGATFIVPKKLGFSLAQEILLNAGSYRGADLEKRGV
TmpE                        SDIVLLSKESYYTLNFMKYGFTPGMGSTLIVPTKLGTDLGNEMLLSANQYRGDELQQRGV
TaY                         ADFVVLSRESVYTTNFMRYGFTPGMGATYIVPKRLGYSLGHELLLNARNYRGADLEKRGV
OocC                        ADFVIMSKESIYTANFMKYGFTPGMGGTLIVPEKLGLALGHEMLMLGNNYRGEALRQRGV
CurF                        ADFVVFSQESVYATNFMKYGFTPVGATSLILREKLGSELAQEMIYTGENYRGKELAERGI
JamJ_ECH_1_14-255           ADFLVLSRESFYANNLMKYGLTPVGSTSLILPKKLGAELGQEMIYTGRQYRGEELAQRGI
Bat3_ECH_1_2986-3226        SDFVVFSQQGCYESNYMKFGFTPGAGATLIFPEKFSSGLAHEILFTGRQFTGSDLEQRGI
PedI_ECH2_2878-3125         CDCSVYSEESIYQSPYMRYDFTPGAGSTLIFPLRLGHDLSREVLLTAREFHGRELRRRGI
OnnB_ECH_2_2961-3192        CDQVIFSLESLYQSPYMQFGFTPGAGSTLIFPHRFGKDFAGEILLSASAYRGSDFERQGI
NspA_ECH_2_3211-3455        CDVNIFSEESIYQTPYMQYGFTPGAGSTLIFPARLGPNLSREILFGGREYKGAELKERGI
DipP_ECH_1_2970-3215        CDYTIFSEESVYQSPYMRYGFTPGAGSTLIFPQRFGKILSREILFTASEFKGKELKERGI
PsyA_ECH_1_1536-1762        CDVTIMAEERVYEAPYMRYGFTPGAGSTLVFPARLGFDLAGEILFTGESYRGAELRDRGI
PhmE_ECH_2_2206-2441        CDQAIFSEESLYFSPYMQYGFTPGAGATLIFPDRFGWDLGREILFTAREYKGHELRSRGI
PhmI_ECH_2_1102-1349        CDGAVFSAESVYHSPYMQYGFTPGAGSTLIIPHQFGHDLGGEILFGAREFKGHELKRRGI
Fr9GH_ECH_2_3120-3369       CDAVLFAEESVYHSPYLSYGFTPGAGSTLVFPMRLGLDLGREILFGAQPYKGRALRERLP
OocJ_mECH_2_2220-2467       CDEAIFSLESVYYSPYMRYGFTPGAGATLILPSRLGLDLAREVLFTANEYKGYELHARGL
                            .*  : : .  *    : :.:**  . : :.  ::.  :. *::  .  . *  :  :
```

```
BatE                        PFP-VLPRAEVLEYALQLARDLAE-KPRISLVTLKSHLVAEMRARLPTIVEQEIAMHDKT
MupK                        PFA-VLPRAQVLAHALELAMELAD-KPRLSLVTLKHHLNRKLREDLPAVIERELAMHALT
CalR                        PFT-VLPRAEVLPEAMKMSRQLAE-KPRVSLMTLKDHLVADQRRKLPDIIRSEVAMHEQT
BaeI                        PFP-VLPRQEVLDYAQELARELAE-KPRNSLITLKRHLVTPFREQLPAVIEKELIMHDET
VirE                        PVR-VVAHDDVASTAHALAAGVAN-APRASLELLKQDLAAPLLAATDAAIEREAAMHRTT
pksI                        PFK-VLPRAEVLDYAVELAQELAE-KPRNSLVTLKDHLVAPLRDQLPRVIEQELMMHEKT
TmpE                        PLQ-VLLRKQVLSRAYELADLIAQ-KPIRSVHLLKQHLCQNICTQLPKYIEQELAMHAQT
TaY                         PFP-VLPRKEVLPHAYEIARDLAA-KPRLSLVTLKRHLVRDIRRELPDVIERELEMHGIT
OocC                        PFP-VYPRADVAAAALSIARDLVK-KPRLSLITLKAHLTQHLRAQLSTYTAQEVLMHEKT
CurF                        PFP-VVSRQDVLNYAQQLGQKIAK-SPRLSLVALKQHLSADIKAKFPEAIKKELEIHQVT
JamJ_ECH_1_14-255           PFP-VLPRKEVLNSARKLAKEIAE-KPRLSLLALKENLTAEIRTKLPEVVSKELEMHEKT
Bat3_ECH_1_2986-3226        PYP-VVHRDQVLVQAIDLARSLCQ-SPRIALMELKEHTANPIREKLQRTYEKEVAMHQKT
PedI_ECH2_2878-3125         DMP-VLPRRQVLAYALSLAAHLAI-TSRQQLVEQKNRRVQPVRELLPRFFEQELAMHEKT
OnnB_ECH_2_2961-3192        QMP-VLPRRRVLAHAMEMANEMAL-SSRDELVETKVVRCQHLRHRLEETYLQELAMHEKT
NspA_ECH_2_3211-3455        SMP-VLPRSQVLGYALKLAHHLAK-FSQNELIQRKNQATAGLRDLLDKVHTQELAMHEKT
DipP_ECH_1_2970-3215        SMP-VLPRKQVLFYSLKLAKKLSL-LSKKELELQKKYYSHSLQKKLSTIFLHELTMHDYT
PsyA_ECH_1_1536-1762        AWP-VLPAGEVETYALQVAREWAR-RPREVLMAEKAARCTPLRDRLADTLAQELSMHEAT
PhmE_ECH_2_2206-2441        NMDVVLPRAQVLGQALQRAGELAR-QSRHELVRRKAERSRLLRTLLEPNYAQELAMHDKT
PhmI_ECH_2_1102-1349        TMP-VLPRSDVLAGAMAAAHHLAS-ASRQTLMDWKQTRVRVLRQRLESIYAAELAMHEQT
Fr9GH_ECH_2_3120-3369       GLS-VVPRGEVLAQARRLAAQWAMRRTRDALIRDKGRRLAALAQALPAMIRKELDMHELT
OocJ_mECH_2_2220-2467       KNP-VLERARVLPFALAVANALAL-QPRAELVAAKTQRSQRLRDQVEANYERELQMHERT
                                *     *   :   .          :   *                *  :*  *

BatE                        FHQ-PEVKARIENVFGN------
MupK                        MHQ-PEVGERISTLFGR------
CalR                        FHQ-DEVKQNILSLFGK------
BaeI                        FHD-DEVKKRITTMYGV------
VirE                        FRL-PEVMDRIVDGYGAAGLRHP
pksI                        FHH-EEVKSRIKGLYGN------
TmpE                        MHQ-PEVANNIKQLFGQ------
TaY                         FHH-DDVRRRIEQLFL-------
OocC                        FFQ-PEVRQNIDRLFGN------
CurF                        FNQ-PEIASRIQQEF--------
JamJ_ECH_1_14-255           FQQ-VKVASQ-------------
Bat3_ECH_1_2986-3226        FVDNPEVLAR-------------
PedI_ECH2_2878-3125         FVGNQQVIEKLNRHFY-------
OnnB_ECH_2_2961-3192        YVGNPDVFAKIQ-----------
NspA_ECH_2_3211-3455        FVGQSQVLKNIQERF--------
DipP_ECH_1_2970-3215        FVNNPSVITNIERYF--------
PsyA_ECH_1_1536-1762        FVGRDDIAAAIEASF--------
PhmE_ECH_2_2206-2441        FVGNQEVQERIRKYY--------
PhmI_ECH_2_1102-1349        FVGSQEVRARIETLYGG------
Fr9GH_ECH_2_3120-3369       FRHNPDVARLIDRAYG-------
OocJ_mECH_2_2220-2467       FVGNQEVLSRIQQHF--------
                                 .:
```

MacpD/TacpD/DifC

```
DifC            MEQTKEM-KQIIAQIIQDIQEQQSYRA--VEAGDDVRVIEDLGFSSLDIA
MacpD           MLNHQVM-DQVFDQVEHQIAQVLGAKGGPLVAVEIDSRFSDLGLSSLDLA
TacpD           MLNVQQIKNEVLDSIIQEVATILSEKLLPITHIENSTAFSELSLTSLDLA
                * : : : .::: .: :::    . :   :   :    :.:*.::***:*

DifC            QLVAQMEMETGVDPFSQGETISSITTVGSICDIYQKY-------------
MacpD           TLISNLEAVYGTDPFADAVAITSIVTVADLCRAYAQQG--VPGPSPDPLD
TacpD           ELISNLEARYEVDPFEELVAITSIVTIEDLASAYALSLSGNTEDSHDLLS
                 *:::::*    .*** :  :*:**.*: .:.  *

DifC            -----MDSAQS
MacpD           AQ--LRDLRQL
TacpD           DELKAIKNQVR
```

MupQ/TmlQ/DifD

```
DifD            MKTESGGDISWVTDKLRETQDSEFLIFKGQTFTYDDLLARISFFQKELEQ
MupQ            MRE----ERNWISDRLASFGDGIALIADGQQSSFNAIGQAMMSWQALLAH
TmlQ            MN------RTWISDRVNSFGDSLAVVEGDNQTTFNAITKSTASW---LEN
                *.       .*::*:: .  *.  ::  .:  ::: :      :   * :

DifD            NGIREGECVALIGDYSPNAIFLLMALLLNRNIVVPLSNESREKHLDMFED
MupQ            QNIACCT-VIINGDYSCAMVGAILALYTRHCVVVPLNRSSSERCAAVART
TmlQ            ESIAPQSVIVINGIFNATSIGLILCAYIKECIIVPLFNYDEEKVRTILDM
                :.*     : : * :.   :  ::.   .. ::*** . . *:   :

DifD            AKVNQIIEINENDSWTLRGGSESASHP-LLDDLRERGESGIIIFTSGTSG
MupQ            TGASWIIDADHSGDLQAL-TTQLPTHE-LVAGLFDRREAGLILLSSGSSG
TmlQ            VQPDLLCDTRSDTARCILFDTPPNSKPALFREINTRRESGLVLYSSGSTG
                .  . : :   .        :   ::  *.  :  * *:*::: :**::*

DifD            KSKASVLSAGRLLEKYKTAKR-KPLRTLIFLKLDHIGGINTLFAILFNGG
MupQ            EPKAILLSLERLFAKYRDAPRSRPRTTAAFLLLDHIGGFNTLLHCLFSGS
TmlQ            TPKAILLSLDKLLHRYQQHNQRSALTIAGFLFFDHIGGFDVMMQCLLTGN
                 .** :**  :*: :*:   :  .     ** :*****::.::  *:.*.

DifD            TIVTSDSRTPESVYQAVDKHAVQVLPATPTFLNMLLMSKADAGYNLSSLK
MupQ            TLVKLDSRDAVSICRQIAEHRIQLLPATPTFLNMLLWGRAYETYDLSSLE
TmlQ            TLVSMTQRTPDDVCKAIEKHRINVLPTTPTFLNMLLINRAYQRCDLSSLT
                *:*.  .* . .: : : :* :::**:********* .:*    :****

DifD            LITYGTEPMPQSTLRGIHRLFPDVRLKQTYGLTELGIFSTKSKDSQSTWM
MupQ            LVTYGTETMPESTLAALNEVFPAVRFKQTYGMSELGILSTRSESSDSLWI
TmlQ            VIAYGSEVMPKATLNLLHQALPEVMLKQTYGMSELGVLPTESKMGNSLWL
                :::**:* **::**  ::. :* * :*****::***::.*.*: .:* *:

DifD            KVGGAGTETKITGGTLWVRSESAMLGYLNAPSPFDEDGWYDTGDQVETDG
MupQ            QIATDDVQVKVVDQILWIKTRTAMLGYLNAPSVIDQEGWLCTGDLVETRG
TmlQ            KIKKA--KYKVQDGVLWIKSPTAMLGYLNKESDELNDEWLCTGDLVEEQG
                ::     : *: .  **::: :*******  *   :: *  *** **  *

DifD            EYIRILGRKSEIINVGGEKVFPAEVESVFLEIPNVRDVLITGRKNPITGE
MupQ            DYFRILGRGESLINVGGLKVLPVEVESRLLSLPFVKDAVVWGRKSPVTGQ
TmlQ            EYIRILGRQSTVINVAGEKVFPAEIEALLLQIPYVKNSLVWGKKSHITGK
                :*:***** . :***.* **:*.*:*: :*.:* *:: :: *:*. :**:
```

```
DifD            IVAAEAVLEKEEDIKSFRKRAVAHCKQKLQPYQIPQIISVTDRKLSNDRF
MupQ            IVAATVVLEGELDTDDTKRQIMLHCQQGLEDFKVPRHIEFVTGRLHSDRF
TmlQ            IVAATIFIDEDIDQKQAKKHISDFCKQALEPYKIPRYFEFVNDPYHSERF
                ****  .:: : * .. :::   .*:* *: :::*: :...     .:**

DifD            KKIR--NV
MupQ            KKIKVASQ
TmlQ            KKIN--KI
                ***.  .
```

MupS/TmlS/DifE

```
DifE            MNNG-----------------KVAVVSGGSRGLGKAIVQTLLDEDYKVAA
MupS            MTDAVSDALHTRHDSAPSAVKGTIIVSGGSQGLGLTTVRCFLEAGYNVAT
TmlS            ME----------------SENGTVIISGGSRGLGFEIASQFLAQGYAVAT
                *                    . ::****:***   .  :*  .* **:

DifE            FSRSESDFINTLRESEKYRGRFYWDAADAADDAAMKQFVLQVYRKFSRID
MupS            FSRRESPAVTELSE----RADFHWQALDCTDYSALTAFVQQVEKRFGGLD
TmlS            FSRGSSEQVVALAE----EPRFFWKSVDGSDYQALTEFLKEAQKKLGNIV
                *** .*  :  * *    .  *.*.: * :*  *:. *: :. :::. :

DifE            GLVNNAGLNLDQLLPLTNDEDIDRILNLNIGSVIKLTRNVSRVMLKQNSG
MupS            GLVNNAATGVEGILSTMRVADIDSALDINLKGQLYLTKLVTAKLLKRGAG
TmlS            GLVNNAAIGADGVLSTMRTSDIDRAIDVNLKAQIYLSKLVSKKLLQNRDG
                ******. . : :*.  .  ***  :::*: . : *:: *:  :*:.  *

DifE            SIVNISSIIGSRGFKGTSVYSASKAALDGLTRSLARELGSKGIRVNSLAP
MupS            SVVNVSSINALRGHSGLTVYSATKAAMDGLTRSLAKELGPRGIRVNSVSP
TmlS            FIINISSIMGVRGLPGVSIYSATKAAMDGITRSLAKELGRKGIRVNSVSP
                 ::*:*** . **  * ::***:***:**:*****:*** :******::*

DifE            GFVDTDMTKNMPEKQKSQIIRRTPMGRLGETDDMTGLVRFLLSPESSFMT
MupS            GYFSSDMVKDLSPQTLSRIERRTPLGRLGTQQEVADLILYLV-DRGTFVT
TmlS            GYFSSDMVKDLSDDILRKIERRTPLGRLGTQDEIAKLVLFLA-TEGKFIT
                *:..:**.*::. .   :* ****:****  :::: *: :*   ...*:*

DifE            GQTMAIDGGLTC
MupS            GQNIAVDGGFTC
TmlS            GQNIVIDGGFTC
                **.:.:***:**
```